

Essays on Migration and Intergenerational Mobility

Jan Stuhler

Thesis submitted to the Department of Economics
in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

University College London

London

August, 2014

Declaration

I, Jan Stuhler, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Jan Stuhler

Abstract

My dissertation addresses questions in two topic areas, intergenerational mobility and migration. I first study the dynamic response of income mobility to structural changes in a model of intergenerational transmission. I illustrate that mobility today depends on past policies and institutions, such that a decline in mobility may reflect past gains rather than a recent deterioration of “equality of opportunity”. How to measure income mobility is addressed in the next chapter, in which I document that heterogeneity in the shape of income profiles generates large life-cycle biases, which cannot be eliminated with standard methods used in the current literature. Finally, in the fourth chapter I study how elements of the transmission process affect the relation between mobility over two generations and the long-run persistence of economic status within families. I provide various arguments why long-run mobility is likely lower than predictions from intergenerational evidence suggest.

In the final chapter I analyse the effect of migration on labor markets. Triggered by the fall of the iron curtain, Germany experienced a sudden inflow of Czech workers that reached a local employment share of up to ten percent. I exploit this natural experiment to assess how immigration affects native workers, and to examine the mechanisms by which labor markets adjust. I find a strong and rapid response in both native wages and employment, and document substantial heterogeneity across age groups: native employment decreases most strongly among older workers, even though their wages are less affected than for other age groups. This finding suggests that the elasticity of labor supply differs across demographic groups, with important implications for the analysis of responses to labor supply shocks. When distinguishing between the different types of adjustment, I find that native employment decreases predominantly through diminished inflows into work, less so through outflows into non-employment.

Contents

1	Introduction and Summary	11
1.1	The Immigration Equation	11
1.2	Intergenerational Mobility	16
2	Interpreting Trends in Intergenerational Mobility	19
2.1	The Literature	22
2.2	A Model of Intergenerational Transmission	23
2.2.1	The Importance of Past Transmission Mechanisms	27
2.2.2	From Simple Examples to Non-Monotonic Trends	28
2.2.3	Intergenerational Mobility in Times of Change	33
2.3	Empirical Application	37
2.3.1	The Swedish Compulsory School Reform	38
2.3.2	Compulsory Schooling in the Intergenerational Model	39
2.3.3	Data	41
2.3.4	Empirical Evidence	43
2.4	From Generations to Cohorts	51
2.5	Conclusions	55
3	Life-Cycle Bias in Intergenerational Mobility Estimation	76
3.1	The Intergenerational Mobility Literature	79
3.2	Measuring Income at a Certain Age	82
3.3	Empirical Evidence on Life-Cycle Bias	85
3.3.1	Data Sources and Sample Selection	85
3.3.2	Empirical Strategy	87
3.3.3	Empirical Results	88
3.3.4	Robustness tests	91

3.4	Extensions	93
3.4.1	Extending the Generalized Errors-in-Variables Model	93
3.4.2	Multi-Year Averages of Current Income	96
3.5	Conclusions	97
4	Mobility Across Multiple Generations:	
	The Iterated Regression Fallacy	108
4.1	The Iterated Regression Fallacy	111
4.2	Models of Inter- and Multigenerational Transmission	112
4.2.1	Indirect Transmission	114
4.2.2	Identification from Multigenerational Data	115
4.2.3	An Additional Factor	116
4.2.4	An Additional Generation	118
4.3	Conclusions	123
5	The Impact of Immigration on Local Labor Markets:	
	Evidence from the Opening of the German-Czech Border	127
5.1	Introduction	128
5.2	An Equilibrium Model with Heterogeneous Labor Supply	132
5.3	Background and Data	139
5.4	Empirical Strategy	144
5.5	Results	149
5.6	Discussion and Conclusions	162

List of Tables

2.1	Sample Statistics by Birth Cohort	58
2.2	Reform Effect on Educational and Income Mobility, Cohorts 1943-1955 . .	59
2.3	Reform Effect on Educational and Income Mobility, Cohorts 1966-1972 . .	60
2.4	Robustness Tests	60
3.1	Summary Statistics by Birth Year of Sons	100
3.2	OLS Estimates of Elasticities and Life-Cycle Bias	100
3.3	Decomposition of Life-Cycle Bias	101
3.4	Correlations Between Residuals and Characteristics	101
3.5	Summary of Robustness Tests	102
3.6	Summary of Cohort Differences, Averages over Ages 31-35	102
4.1	Inter- and Multigenerational Persistence in Educational Attainment . . .	125
4.2	The Grandparent Coefficient in Educational Persistence	126
5.1	Characteristics of Treated, Inland and Matched Control Districts	171
5.2	Characteristics of Czech and Non-Czech Nationals in Border Region . . .	172
5.3	First Stage	172
5.4	Wage and Employment Baseline Estimates by Skill	173
5.5	Robustness Checks	174
5.6	Employment and Wage Effects by Skill and Age Groups	175
5.7	Margins of Adjustment: Unemployment versus Movements between Areas	176
5.8	Inflows vs. Outflows	177

List of Figures

2.2	A Declining Impact of Parental Income and Increasing Returns to Skills .	61
2.1	A Change in the Heritability of, or Returns to, Endowments	61
2.3	A Swap in Prices	62
2.4	Raising the Compulsory Schooling Level	62
2.5	Share of Offspring and Fathers Subject to Reform	63
2.6	Mean and Variance of Years of Schooling over Cohorts	63
2.7	Trends in the Intergenerational Educational Coefficient over Cohorts . .	64
2.8	Educational Attainment and Mobility, Pre- vs. Post-Reform	65
2.9	Heterogeneity in the Reform Effect over Cohorts	66
2.10	Trends in the Intergenerational Educational Correlation over Cohorts . .	66
2.11	Placebo Test: Second Generation	67
2.12	Trends in Conditional Intergenerational Regression and Correlation Coef- ficients	68
3.1	Illustrative Example of Log Annual Income Trajectories	103
3.2	OLS Estimates of $\lambda_{s,t}$	103
3.3	OLS Estimates of Elasticities and Life-Cycle Bias	104
3.4	Estimates of Life-Cycle Bias for Different Age Spans (Cohort 1955-57) . .	104
3.5	OLS Estimates of Elasticities for Various Cohorts	105
3.6	OLS Estimates of Elasticities with Right-Side Measurement Error	105
3.7	OLS Estimates of Elasticities with Both-Side Measurement Error	106
3.8	OLS Estimates of Coefficients in Standard and Extended GEiV Model . .	106
3.9	Remaining Life-Cycle Bias in Standard and Extended GEiV Model	107
3.10	Remaining Life-Cycle Bias in GEiV Model with Multi-Year Averages . . .	107
5.1	Border Region	178

5.2	Employment Shares of Czech nationals: Border vs Inland	179
5.3	Spatial Distribution of Czech Nationals in Border Region	179
5.4	Aggregate Wage and Employment Effects	180
5.5	Synthetic Control Method, Wage and Employment Effects	181
5.6	Yearly Native Inflow and Outflow Effects	182

Acknowledgements

Funding arrangements

The writing of this thesis was supported by scholarships from the German National Academic Foundation (*Studienstiftung des Deutschen Volkes*) and University College London (*David Pearce* and *Impact* Scholarship), and by grants from the Centre of Research and Analysis of Migration and the NORFACE Migration research project.

Joint work

Chapters 2 and 3 are based on joint work conducted with Martin Nybom, Swedish Institute for Social Research (SOFI). Chapter 5 is based on joint work conducted with Uta Schönberg and Christian Dustmann, University College London (UCL). The remaining chapters are my own work, as is responsibility for any errors.

Helena Holmlund (IFAU and SOFI) kindly provided her program files on the reform coding that I used in chapter 2. I am further grateful to members of the Institute for Employment Research in Nuremberg, which provides the data used in chapter 5 of this thesis; in particular Marco Hafner, who provided much help during the early stages of my work. Finally, I received friendly help from various IT staff and other researchers at UCL and the Institute for Social Research in Stockholm, notably from John McGlynn, Fatima Cetin, and Kristian Koerselman.

Personal acknowledgements

I rather enjoyed writing this thesis – at least, most of the time. This is an unusually positive assessment, and I am grateful to those people who lightened the journey and helped to overcome the difficult parts.

First, I am grateful to my supervisors, Christian Dustmann and Uta Schönberg. Their ideas, experience and knowledge pushed our project on the effect of immigration on labor markets – on which I spent the majority of my time – much further than what I would have thought possible. This process was not without conflict. But I profited tremendously from their advice, and I am grateful for their strong support through all the stages of my dissertation, in particular the important job market phase.

My third co-author is Martin Nybom, with whom I collaborated on the first two chapters. Meeting Martin during the first year of my dissertation, at a student conference in Toulouse, was a stroke of luck that greatly formed this thesis, and a decisive factor why I actually enjoyed writing it. I profited in particular from Martin’s efficiency and ability to “sort the sheep from the goats” – illustrated by the fact that not only did he finish his dissertation years before me, he also cared for two children in the meantime.

Martin introduced me to Anders Björklund, Markus Jäntti, Matthew Lindquist and others from the Swedish Institute of Social Research in Stockholm, who lend considerable and generous support not only to our joint, but also my own work on intergenerational mobility. Anders, Markus and Matthew’s friendly “adoption” led to several visits to the institute in Stockholm, whose busy and friendly atmosphere I enjoyed tremendously.

Less direct but equally important was the contribution of friends and colleagues at University College London and the Centre for Research and Analysis of Immigration: Anna Raute, with whom I shared office and else in the first years; Luigi Minale and his charming comments; Ilker Kandemir, with whom I had the most extraordinary discussions; Italo Lopez and Christopher Rauh, with whom the job market was substantial more fun; and Marco Alfano, who made crucial contributions in the categories food and fitness. I fondly remember visits all over North-East London and the Polish countryside, with Thomas Cornelissen, Ines Helm, Anna Okatenko, Simon Gorlach, Anna Rosso, Florian Oswald, Thibaut Lamadon, my numerous flatmates, and many others: they are too numerous to mention, and I am afraid to forget someone.

Finally, there a number of people as of whom I decided to pursue a dissertation in

the first place. Benny Moldovanu gave a superb lecture on matching markets at the University of Bonn that left lasting impression. My fascination for microeconomic theory however lost eventually ground against my interest for empirical labor. Partly at fault are all the friendly people who I met during my time as research assistant at the Institute for the Study of Labor, in particular Holger Bonin, Thomas Dohmen, Mark Fallak, Holger Hinte, Hilmar Schneider and Uwe Sunde.

An important experience was also my year abroad at the University of California in Berkeley, in particular the classes given by George Akerlof, David Romer and Enrico Moretti. I made a great many of friends, among them an exceptional group of economists: Bjarne Steffen, Martin Hackmann, Juan Carlos Suarez and Marieke Schnabel – with whom I happened to share every single station of my studies in Bonn, Berkeley, London and Stockholm. Finally, David Jaeger gave a fantastic course on labor economics during my final year in Bonn. His enthusiastic support for my (delayed) applications to graduate school was invaluable at that otherwise difficult time.

These words suggest that teaching matters, and I also wish to thank Donald Verry and Toru Kitagawa from UCL, for whom I enjoyed working as a Teaching Assistant – both are great teachers, whose support extends far beyond the classroom (Donald’s fame in this respect is already cast in stone).

Finally, I would like to thank my family, for providing the support that is necessary to embark on something as challenging and peripheral as a PhD in Economics; in particular my parents, Brigitte and Ulrich, who worked very hard to ensure that difficulties and setbacks at home had little consequences on my studies. This thesis is dedicated to them; my siblings, Simon and Anne, who might even read it; and of course my partner Alícia Escribà, who throughout showed a very healthy attitude – mild curiosity about its content, a more pronounced interest in when it will be finished, and the assurance that other important things in life await if the dissertation turns sour. It did not, and is finally finished!

Chapter 1

Introduction and Summary

This thesis consists of four self-contained essays, on *intergenerational mobility of economic status* and *the labor market effect of migration*. I spent about half of my work on each topic, resulting in three chapters of different length on intergenerational mobility (the first two are based on joint work with Martin Nybom, Swedish Institute for Social Research) and one long chapter on the impact of migration on local labor markets (joint with Christian Dustmann and Uta Schöneberg, University College London). I start with a more detailed introduction of my work on migration, and provide then a short summary of the other three chapters.

1.1 The Immigration Equation

Sometimes one strikes it lucky, as in my project on migration. It started in 2009 as a collaboration with my supervisors Christian Dustmann and Uta Schöneberg, who tasked me to study how the fall of the iron curtain and German reunification affected labor markets in Germany's border regions. Provided with data from German social security records, my mission turned quickly interesting when I discovered that in the early 1990s, a high share of workers in Eastern Bavaria were of Czech nationality. It was easy to understand why Czechs at that time *wanted* to work in Germany, where the wage level was multiple times higher than in the Czech Republic. But I was surprised that they were *allowed* to enter, as the movement of labor from Central and Eastern European countries to Germany was (and still is) heavily restricted. Moreover, the distribution of those Czech workers was odd – their inflow occurred very suddenly, and while they were heavily concentrated in rural municipalities close to the Czech-German border, their

share in other areas was negligible.

A search in legislative texts revealed the source of that inflow – a policy that gave permission to Czech and Polish workers to work, but not to live in the German border region. The commuting policy received little attention in the national press at the time, as it applied only to a small number of remote areas. The “*Grenzgängerregelung*”, which existed in similar forms along the Western and Southern German borders, was in turn possible only because of the unexpected revolutions of 1989 that preceded its introduction. As such, we observed a sudden, rapid, unanticipated, and large labor supply shock, triggered by political events that did not relate to the specific conditions in the German border region. Moreover, the strong variation of the share of Czech workers within Germany – commuting is costly, which explains their low share in municipalities farther from the border – in an instrumental variable strategy. Finally we had access to population-wide panel data of all men and women covered by the social security system in Germany, which allowed us to not only assess labor market outcomes in detailed subgroups, but also to follow affected individuals over time and space.

A short discussion of the existing literature will illustrate why these conditions were so valuable. The literature on the labor market effects of immigration on labor markets is large, but little agreement exists on the degree to which it can have adverse effects on the employment and wages of native workers.¹ The literature lacks consensus even while it features some of the most elegant and influential papers in economics: Card’s classic study on the impact of the Mariel Boatlift (Card 1990), in which a large number of Cubans entered the Miami labor market, is one of those papers that attracted me (and probably many others) to the field of labor economics in the first place.

However, the literature faces countless problems. Immigrants tend to self-select into areas in which the local economy is growing, giving rise to a positive spurious correlation between immigrant inflows and native outcomes. Second, the local impact of immigration might progressively dissipate to the rest of the economy as capital, firms and the native population respond. Over time it becomes therefore increasingly unclear how the local effect of immigration relates to its overall impact on the host economy. Third, the tendency of immigrant inflows to cluster in specific dimensions (e.g. location, occupation, or education) complicates the statistical analysis. Labour market conditions vary along

¹An entertaining account of that lack of consensus and the main contributors to the literature is given by Roger Lowenstein in his article “The Immigration Equation”, published in the New York Times on July 9, 2006.

those dimensions for other reasons than migration. This needs to be taken into account for robust statistical inference, which is not possible in simple before- and after-type of analyses, as are often employed in the literature.

In particular the first two problems are at the center of attention in the literature, and have been addressed in various ways. The spatial or *area approach* exploits differences in immigrant density across areas, but has been criticised with the argument that immigrant inflows might lead to native outflows, dissipating their effects across the wider economy. The *skill-cell approach* instead exploits differences in immigrant density across skill and demographic groups – often under strong assumptions, which in this chapter 5 I show to not always hold.

Given the amount of existing work in the literature, and the difficulties it faces, an observer may be excused for concluding that we should just give up to find any consensus. But not only is the provision of factual information directly relevant for public policy in this politically sensitive area. The core questions in the immigration literature are essentially questions on how labor markets work, and the idea that changes in the supply factors of production generate responses in its prices is at the core of economic theory. If we do not agree how natives are affected by immigration-induced labor supply shocks then perhaps we do not understand how labor markets work in general.

In chapter 5 we revisit the “immigration equation”, by studying the consequences of the commuting policy in detailed, population-wide panel data on individual workers. Studying the short-run response, we document a modest but rapid decline in wages, and a large decline in the employment of native workers. The observed response supports one of the core hypothesis held by the literature: as of imperfect substitutability, the negative effect of immigration falls most strongly on those workers whose qualifications become relatively more abundant. In our setting, the inflow of predominantly unskilled Czech workers led to particularly negative responses in employment and wages for unskilled native workers.

However, our findings are in stark contrast to other assumptions that are standard in the literature. First, we find both skilled and unskilled workers to be negatively affected – in contrast to the strand of literature that assume immigration to have (or that can identify) only distributionary, but no average effects. More importantly, we find that the notion of imperfect substitutability, while useful, does not suffice to explain the distributionary effect of immigration. In particular, we find the decline in employment to

be strongest among older native workers, even though only few Czech commuters enter that age group.

Within our theoretical model, we can explain that finding by variation in workers' responsiveness to wage changes, that is, heterogeneity in the elasticity of labor supply across groups. This innovation, while consistent with evidence from other literatures, overturns standard assumptions on the distribution of the wage effect of immigration – with potentially severe implications, as these assumptions are directly used for the identification in one of the two dominant strands of the literature, the skill-cell approach. Our results suggest that while differences in immigrant density across education groups can indeed be exploited, extension of that strategy along other dimensions, such as age or experience, are fragile, as workers are likely to respond differently to wage changes along those dimensions.

Overall, our results appear to paint a quite negative picture on the effect of immigration on native labor market outcomes. However, the results are less dramatic upon closer inspection. First, we focused on the short run, and the region under consideration is rural and received only few migrants before the fall of the iron curtain. Such areas may be less able to absorb large immigrant inflows than, for example, Miami or California. The particularities of the observed supply shock are also important. The unexpected and sudden arrival of a large number of migrants is likely to have far more negative effects than the gradual and potentially anticipated arrival of smaller numbers. Second, Czech workers had to commute, such that their contribution to local consumption was comparatively small. Indeed, we find that local population levels decreased in response to Czech inflows. The absence of a simultaneous demand shock distinguishes our setting from other studies in the literature.

As such, the numbers presented in this study cannot be directly applied to other immigration-induced supply shocks. Instead, our estimates reflect the result from what I consider an atypical, but highly informative experiment – an experiment that in my view provides a more direct test of the core theoretical models in the literature than previous empirical work. Theoretical work almost always considers sudden, unexpected supply shocks that occur in isolation – ignoring the fact that immigrants may also shift local demand. Our “pure” supply shock comes closer to that stylised thought experiment than other natural experiments that have been studied in the literature. Our results suggest that such shock can indeed have substantially negative, both absolute and distributionary,

effects on native workers. This result may come as a relief to those who argue that standard theories of factor demand are simply not consistent with studies that find immigration to have no negative effects on natives even in the short run.

Our findings may thus help to reconcile some of the existing arguments in the literature. However, the availability of population-wide of our data allowed us also to extend our work in various new directions, and to study aspects of the adjustment process in local labor markets about which only little is known – the mechanisms via which the level of employment adjusts, the timing of those processes, and the degree to which they matter for different groups of workers.

First, we find the decline in native employment to be predominantly explained by movements from and to non-employment, while geographic movements across areas are far less important. The relative importance of the two channels varies with education and age, with young or skilled natives being more geographically mobile than older and unskilled workers.

Second, we decompose the overall employment effect further into changes in inflows – native workers who enter local employment – and changes in outflows of existing workers. Our findings suggest that inflows explain a far greater share of the total employment response than outflows, in stark contrast to the way employment responses to immigration are commonly interpreted. Inflows respond also more rapidly than outflows, which may explain why employment can adjust so rapidly – the adverse effect on native employment achieves full strength only one year after the inflow of Czech workers reached its peak. Finally, we find the relative importance of the various inflow and outflow channels to differ in distinct ways with age and skill.

These findings are interesting not only for the immigration literature, but more generally, as they shed light on the channels via which labor markets, and native workers, adjust to local shocks. Our analysis provides a coherent picture of these processes, but I believe more work is needed to assess their implications. For example, as each process affects different groups of workers, the welfare implications of immigration may depend crucially on what channel dominates in the adjustment of employment levels. Moreover, it seems likely that the importance of some processes depends strongly on the institutional environment, such as employment protection laws. The ways via which labor markets adjust to immigration, and the type of workers that are predominantly affected, might then be quite case-specific.

1.2 Intergenerational Mobility

The other three chapters shed light on various questions on *intergenerational mobility* – the degree to which status differences are (not) transmitted from parents to their children. My work addresses in particular conceptual issues that relate to the *measurement* and *interpretation* of mobility.

The third chapter in this thesis, based on joint work with Martin Nybom, is a detailed treatment of issues that arise in the measurement of income mobility. Researchers typically observe incomes only over a few years, but the use of such snapshots will cause a so-called *life-cycle bias* if the snapshots cannot mimic lifetime outcomes. We use uniquely long series of Swedish income data to assess how large this bias is in the *intergenerational income elasticity*, the most commonly used statistic in economic research on mobility. We confirm various existing arguments from the literature, in particular that life-cycle bias is smallest when incomes are measured around midlife, a central implication from a widely adopted generalization of the classical errors-in-variables model. However, we also show that the model cannot predict the ideal age of measurement, and that the life-cycle bias can be substantial at other ages. We explain the conceptual reasons behind these empirical findings, and discuss how the generalized model can be extended to reduce life-cycle bias further.

Even if mobility can be correctly measured, how do we interpret changes in mobility? For example, does declining mobility reflect a diminished effectiveness of current policies and institutions in the promotion of “equal opportunities”, as is a common interpretation? In chapter 2, likewise based on joint work with Martin Nybom, we answer such questions. We show theoretically and empirically that mobility trends may instead be caused by events in a more distant past, as structural changes affect mobility over *multiple* generations. We argue that such dynamic responses are of particular importance in the study of intergenerational persistence, since even a single transmission step – one generation – corresponds to a very long time period. Institutional reforms or other systemic changes generate therefore long-lasting mobility trends.

The interpretation of such trends necessitates a dynamic perspective, but existing theoretical work focuses instead on the relationship between causal mechanisms and the implied long-run or *steady-state* level of intergenerational mobility. We thus study the dynamic response of income mobility to structural changes in a simultaneous equation

model of intergenerational transmission. We first show that the level of intergenerational mobility depends not only on contemporaneous transmission mechanisms, but also on the distribution of income and skills in the parent generation – and thus on past mechanisms. This result leads to a number of implications. First, changes in policies and institutions can generate long-lasting mobility trends. Conversely, changes in mobility today might not be explained by recent structural changes, but by major events in the more distant past. Second, differences in mobility across countries, or across groups within countries, might reflect not only the consequences of current but also of past policies, institutions and conditions.

A fairly general class of changes in transmission mechanisms cause *non-monotonic transitions* between steady states, and our analysis suggests that mobility tends to be highest when a structural change occurs. Times of change are thus times of high mobility, while declining mobility today may reflect past gains rather than a recent deterioration of “equality of opportunity”. We finally exploit data over three generations and a compulsory school reform in Sweden to test the dynamic implications of our model. The reform had a large, long-lasting, and non-monotonic effect: it reduced the transmission of disparities in income and education from parents to their offspring in the directly affected generation, but increased intergenerational persistence in the next.

My final chapter on intergenerational mobility relates to both the measurement and interpretation of mobility. While a vast empirical literature has estimated the degree of intergenerational persistence in socio-economic characteristics between parents and their children, there exists little evidence on the degree of *long-run* mobility across *multiple* generations, such as between grandparents and their grandchildren (Lindahl et al, 2014, is a recent exception). In its absence we rely on extrapolations from parent-child correlations, and predictions were routinely derived by exponentiation of intergenerational measures. Importantly, such predictions imply high long-run mobility even when intergenerational mobility is low.

The underlying idea, that regression implies iterated regression, appears quite natural in a linear regression context. However, it turns out to be a common statistical fallacy, which as I discuss in chapter 4 arises frequently also in other economic literatures and disciplines. I then examine how elements of the transmission process affect the relation between intergenerational and multigenerational mobility. Considering direct and indirect transmission, the multiplicity of skills, and the role of grandparents I con-

clude that long-run mobility will likely be lower, possibly much lower, than predictions from intergenerational evidence suggest. I further argue that the observed deceleration of regression to the mean beyond two generations is more plausibly explained by the multiplicity and indirectness of parent-child transmission processes than the existence of important higher order causal effects, such as from grandparents. While my main objective is to provide a theoretical complement to the recent wave of recent empirical studies on the subject, I also provide a brief illustration of my arguments using data on three generations from Swedish registers.

Chapter 2

Interpreting Trends in Intergenerational Mobility

The evolution of inequality in economic status over time is a fundamental topic in the social sciences and in public debate. Two central dimensions of interest are the extent of cross-sectional inequality between individuals and its persistence across generations, as status differences are transmitted from parents to their children. Both have important implications for individual welfare and the functioning of political and economic systems.¹

A significant rise in cross-sectional income inequality from the late 1970s in the US, UK and (more recently) other OECD countries is well documented, but much less is known about trends in *intergenerational mobility*.² Yet, we do know that income mobility differs substantially across countries, and the observation that those differences appear negatively correlated with cross-sectional inequality has received much attention.³ A central theme in the recent literature is thus if income inequality has not only increased, but also become more persistent across generations. This question is debated particularly in countries that experienced rising cross-sectional inequality, such as the US, where commentators argue that low mobility threatens social cohesion and the notion of “American

¹Intergenerational mobility is for example seen to contribute to the stability of liberal democracies, by legitimating income and status inequalities and by reducing the potential for class-based collective action (see [Erikson and Goldthorpe, 1992](#)).

²[Autor and Katz \(1999\)](#) discuss trends in wage inequality across countries. [Atkinson, Piketty, and Saez \(2011\)](#) find a substantial rise in top income shares in the US and various other countries.

³A large empirical literature (see [Solon, 1999](#), and [Black and Devereux, 2011](#)) seeks to quantify how intergenerational mobility differs across countries, groups and time. [Björklund and Jäntti \(2009\)](#), [Blanden \(2011\)](#), and [Corak \(2013\)](#) present evidence on the correlation between cross-sectional inequality and mobility.

exceptionalism”.⁴

But how should evidence on declining mobility be *interpreted* – does it reflect a diminished effectiveness of current policies and institutions in the promotion of “equal opportunities”? In this paper we show theoretically and empirically that mobility trends may instead be caused by events in a more distant past, as structural changes affect mobility over *multiple* generations. We argue that such dynamic responses are of particular importance in the study of intergenerational persistence, since even a single transmission step – one generation – corresponds to a very long time period. Institutional reforms or other systemic changes generate therefore long-lasting mobility trends.

The interpretation of such trends necessitates a dynamic perspective, but existing theoretical work focuses instead on the relationship between causal mechanisms and the implied long-run or *steady-state* level of intergenerational mobility. We thus contribute to the literature by examining the dynamic implications of a simultaneous equations model of intergenerational transmission. We deviate from previous work also by assuming that income depends on human capital through a vector of distinct productive characteristics instead of a single factor. This choice is in accordance with the growing evidence on the importance of distinct, including noncognitive types of skills (e.g., Heckman et al., 2006). We find that such multiplicity also matters in the intergenerational context.

Using our model we first show that the level of intergenerational mobility depends not only on contemporaneous transmission mechanisms, but also on the distribution of income and skills in the parent generation – and thus on past mechanisms. This result leads to a number of implications. First, changes in policies and institutions can generate long-lasting mobility trends. Conversely, changes in mobility today might not be explained by recent structural changes, but by major events in the more distant past. Second, differences in mobility across countries, or across groups within countries, might reflect not only the consequences of current but also of past policies, institutions and conditions.

A fairly general class of changes in transmission mechanisms cause *non-monotonic transitions* between steady states. We show that changes in the relative returns to

⁴Exemplary articles are “*Ever Higher Society, Ever Harder to Ascend*” in *The Economist* (Dec. 2004), “*Moving Up: Challenges to the American Dream*” in the *Wall Street Journal* (May 2005), “*The Mobility Myth*” in *The New Republic* (Feb. 2012), or the recent “*Great Divide*” series on nytimes.com. The political importance of the topic is exemplified by a speech of Alan Krueger, Chairman of the Council of Economic Advisers, who warned that intergenerational mobility should be expected to decline further as of the recent rise in income inequality in the US (speech at the Center for American Progress, January 12th, 2012).

different types of human capital or endowments generate transitional mobility, as some families gain while others lose. Technological, institutional or other structural change may thus increase mobility initially, followed by a decreasing trend that lasts over multiple generations. We conclude that times of change tend to be times of high mobility, while mobility is likely to decrease when the economic environment stabilizes. A shift towards a more meritocratic society has similar consequences. A rise in the importance of own skill relative to parental status is to the advantage of talented offspring from poor families, providing opportunities that were not yet available to their parents. Intergenerational mobility is thus particularly high in the first affected generation, but is bound to decline in subsequent generations. Even structural changes that are clearly mobility-enhancing in the long-run can therefore cause negative trends over some generations.

Declining mobility today may then not signal that current policies and institutions promote equality of opportunity less effectively, but might instead be a repercussion of major *improvements* in the past. These results are important for policy evaluation and for the interpretation of mobility trends. Observed mobility shifts are commonly related to contemporaneous changes in policy or institutions, which may result in misleading conclusions about determinants of the former and long-run consequences of the latter.

A dynamic view of intergenerational transmission does not only reveal such pitfalls, it may also aid our understanding of causal mechanisms (as different structural shocks have different dynamic implications) and of mobility differences across countries and time that have been documented by the empirical literature. Our main objective is to illustrate the general relationship between causal mechanisms and mobility trends, but we comment briefly also on various practical implications that seem particularly relevant for the recent literature.

The paper proceeds as follows. We next discuss the related literature. In Section 2.2 we present our model of intergenerational transmission, derive current and steady-state mobility levels in terms of its structural parameters, and analyze the dynamic content of the model. In Sections 2.2.2 and 2.2.3 we study three theoretical examples to illustrate our main theoretical findings. the introduction of a cohort dimension into our model in Section 2.4. Section 2.5 concludes.

2.1 The Literature

Many studies examine the theoretical relationship between causal transmission mechanisms and the implied long-run or *steady-state* level of intergenerational mobility, but there exists little work on transition paths between those steady states. In the standard simultaneous equations approach as developed by Conlisk (e.g., [Conlisk, 1974a](#)) only [Atkinson and Jenkins \(1984\)](#) focus on systems that are not in steady state.⁵ While they show that failure of the steady-state assumption impedes identification of invariable parameters of the structural model, we instead consider how changes in structural parameters affect mobility in subsequent generations. [Solon \(2004\)](#) notes that the interpretation of mobility trends would benefit from a theoretical perspective, and examines how structural changes (such as in the return to human capital and the progressivity of public investment) affect mobility in the first affected generation. [Davies et al. \(2005\)](#) compare mobility and cross-sectional inequality under private and public education in a model of human capital accumulation. They note that the observation of mobility trends may help to distinguish between alternative causes of rising cross-sectional inequality.

While theoretical work is sparse, it exists much empirical work on mobility trends in the US and other countries. A long-standing and mostly sociological literature is concerned with occupational and class mobility (see [Breen, 2004](#), [Hauser, 2010](#), and [Long and Ferrie, 2013b](#)), examining both absolute (subject to changes in the occupational structure at the aggregate level) and relative mobility rates across countries and time. A more recent but fast-growing economic literature examines mobility trends in *income* or *educational attainment*, which are important indicators and potentially key mechanisms for the reproduction of economic advantage (see [Black and Devereux, 2011](#)). Most economic studies assess how strongly absolute or relative *differences* among parents are transmitted to their offspring, abstracting from mean changes over generations.

Some of the emerging evidence on income mobility appears conflicting, perhaps as a result of the substantial data requirements that such studies face. Measurement ideally requires income data that span over two generations, but often only sparse data are available or exploited.⁶ [Hertz \(2007\)](#) and [Lee and Solon \(2009\)](#) find no evidence of a

⁵Moreover, [Jenkins \(1982\)](#) discusses stability conditions for systems of stochastic linear difference equations with constant coefficients, [Conlisk \(1974b\)](#) derives stability conditions for systems with random coefficients.

⁶In chapter 3 I summarize methodological advances in the recent literature, and argue that these can still not fully eliminate *life-cycle bias* in mobility estimates based on incomplete income data. This bias

major trend across cohorts of sons born 1952-1975 in the US, but cannot reject more gradual changes over time. [Levine and Mazumder \(2007\)](#) as well as [Aaronson and Mazumder \(2008\)](#) argue that mobility has fallen in recent decades – the latter based on intergenerational estimates from synthetic families (constructed from census data), the former based on estimates of sibling correlations in various economic outcomes. Such decline has also been found for the UK, in [Blanden et al. \(2004\)](#) and [Nicoletti and Ermisch \(2007\)](#).⁷ Other studies examine how educational mobility differs between groups, how it is affected by institutional aspects, or how it changes over time. [Hertz et al. \(2008\)](#) present trends in educational mobility over 50 years for 42 countries, noting that Nordic countries display comparatively high intergenerational mobility.

A central concern in many of these papers, policy-related outlets, and the public press is that mobility may have declined in conjunction with the recent rise in income inequality.⁸ Various potential causal factors for observed trends – such as educational expansion, rising returns to education, or changes in welfare policies – are considered in the literature (e.g., [Levine and Mazumder, 2007](#), and further articles in the same issue). Common to all explanations is that they relate trends to recent events that may have directly affected the respective cohorts. We argue that this is only one potential interpretation, and that the key to an understanding of current mobility levels and trends might lie in the more distant past.

2.2 A Model of Intergenerational Transmission

Measuring intergenerational mobility. In our theoretical analysis we consider the intergenerational elasticity of income, which is a popular descriptive measure of persistence in relative economic status. Our main arguments extend to mobility in other outcomes, such as educational attainment, which we will consider in our empirical analysis. Consider a simplified one-parent one-offspring family structure, with $y_{i,t}$ as log lifetime income of the offspring in generation t of family i and $y_{i,t-1}$ as log lifetime income of the parent.

can differ by cohort and may mask gradual changes of mobility, or generate a false impression of such trends.

⁷See [Erikson and Goldthorpe \(2010\)](#) and [Blanden et al. \(2012\)](#) for a debate of divergent findings in measures of income and occupational mobility.

⁸See references in footnote 4 for the US, or [Blanden \(2009\)](#) for the UK.

The intergenerational elasticity is given by the slope coefficient in the linear regression

$$y_{i,t} = \alpha_t + \beta_t y_{i,t-1} + \epsilon_{i,t}. \quad (2.1)$$

The elasticity β_t captures a *statistical* relationship and the error $\epsilon_{i,t}$ is uncorrelated with the regressor by construction. Under stationarity in the variance of $y_{i,t}$ it equals the intergenerational correlation, which adjusts the elasticity for changes in cross-sectional inequality. The intergenerational income elasticity is the most commonly estimated parameter in the empirical literature and captures to what degree percentage differences in parents' incomes tend to be transmitted to the next generation. A low elasticity or correlation indicates high mobility.

A model of intergenerational transmission. We model intergenerational transmission as a system of stochastic linear difference equations, in the tradition of the simultaneous equation approach developed and elaborated by Conlisk (1969, 1974a) and [Atkinson and Jenkins \(1984\)](#). We show in Appendix A.1 that the “mechanical” pathways represented by these equations can be derived from the optimizing behavior of parents in an underlying utility-maximization framework (see [Becker and Tomes, 1979](#), [Goldberger, 1989](#), and [Solon, 2004](#)). The equations of our baseline model are

$$y_{it} = \gamma_{y,t} y_{it-1} + \delta'_t \mathbf{h}_{it} + u_{y,it} \quad (2.2)$$

$$\mathbf{h}_{it} = \gamma_{h,t} y_{it-1} + \Theta_t \mathbf{e}_{it} + \mathbf{u}_{h,it} \quad (2.3)$$

$$\mathbf{e}_{it} = \Lambda_t \mathbf{e}_{it-1} + \mathbf{v}_{it}. \quad (2.4)$$

From equation (2.2), income y_{it} in generation t of family i is determined by *parental income* y_{it-1} , own *human capital* \mathbf{h}_{it} , and *chance* $u_{y,it}$. The parameter $\gamma_{y,t}$ captures a direct effect of parental income that is independent from offspring productivity, which may arise as of nepotism, statistical discrimination under imperfect information on individual productivity, or other reasons.⁹ Human capital consists of a $J \times 1$ vector \mathbf{h}_{it} with elements $h_{1,it}, \dots, h_{J,it}$, reflecting distinct characteristics such as formal schooling, health, and cognitive and non-cognitive skills. These characteristics are valued on the

⁹For example as of credit constraints influencing choices on the labor market, parental information and networks, or (if total market income is considered) returns to bequests. The exact mechanism and the distinction between earnings and income are not central for our purposes.

labor market according to a $J \times 1$ price vector $\boldsymbol{\delta}_t$ with elements $\delta_{1,t}, \dots, \delta_{J,t}$. The random shock term $u_{y,it}$ captures factors that do not relate to parental background. For our analysis it makes no difference if these are interpreted as (labor market) luck or as the impact of other characteristics that are not transmitted within families.

From equation (2.3), human capital \mathbf{h}_{it} is affected by parental income y_{it-1} , own *endowments* \mathbf{e}_{it} , and chance $\mathbf{u}_{h,it}$. A role for parental income may for example stem from parental investment into offspring human capital. Elements in the $J \times 1$ vector $\boldsymbol{\gamma}_{h,t}$ may differ if parental investments are more targeted or more effective on some types of human capital than others. Parental income may thus affect offspring income directly (through $\gamma_{y,t}$) or indirectly (through $\boldsymbol{\gamma}_{h,t}$).¹⁰ The $J \times K$ matrix $\boldsymbol{\Theta}_t$ governs the role that endowments such as abilities or preferences play in the accumulation of different types of human capital. Those endowments, consisting of the $K \times 1$ vector \mathbf{e}_{it} with elements $e_{1,it}, \dots, e_{K,it}$, are partly inherited from parental endowments \mathbf{e}_{it-1} and partly due to chance \mathbf{v}_{it} . The elements of the $K \times K$ matrix $\boldsymbol{\Lambda}_t$ with elements $\lambda_{11,t}, \dots, \lambda_{KK,t}$ govern the *heritability* of each endowment. We consider $\boldsymbol{\Lambda}_t$ to represent a broad concept of intergenerational transmission potentially working through both nature (e.g. genetic inheritance) and nurture (e.g. family environment). The random shock $u_{y,it}$ and elements of $\mathbf{u}_{h,it}$ and \mathbf{v}_{it} are assumed to be uncorrelated with each other and past values of $\{y_{it}, \mathbf{h}_{it}, \mathbf{e}_{it}, u_{y,it}, \mathbf{u}_{h,it}, \mathbf{v}_{it}\}$.

For convenience we drop the individual subscript i and make a few simplifying assumptions. As we focus on relative mobility assume that all variables are measured as trendless indices with constant mean zero (as in Conlisk, 1974a). To avoid case distinctions assume further that those indices measure positive characteristics with a non-negative effect on income (such that $\gamma_{y,t}$ and the elements of $\boldsymbol{\gamma}_{h,t}$ and $\boldsymbol{\delta}'_t \boldsymbol{\Theta}_t$ are non-negative) and that parent and offspring endowments are not negatively correlated (such that elements of $\boldsymbol{\Lambda}_t$ are non-negative), for all t .

Using equation (2.3) to substitute out $\mathbf{h}_{i,t}$ we have

$$y_t = \gamma_t y_{t-1} + \boldsymbol{\rho}'_t \mathbf{e}_t + u_t \quad (2.5)$$

$$\mathbf{e}_t = \boldsymbol{\Lambda}_t \mathbf{e}_{t-1} + \mathbf{v}_t, \quad (2.6)$$

¹⁰The distinction may not be sharp in practice; for example, parental credit constraints might affect educational attainment and human capital acquisition of offspring, but might also affect their career choices for a given level of human capital.

where the parameter $\gamma_t = \gamma_{y,t} + \delta'_t \gamma_{h,t}$ aggregates the direct and indirect effects of parental income, the $1 \times K$ vector $\rho'_t = \delta'_t \Theta_t$ captures the returns to inherited endowments and human capital (affected both by the importance of endowments in the accumulation of and the returns to human capital), and where $u_t = u_{y,t} + \delta'_t u_{h,t}$ aggregates the random shocks in income and human capital.

Our model has a similar structure as the model in [Conlisk \(1974a\)](#), but in contrast to the previous literature we assume that income depends on human capital through a vector of distinct productive characteristics. This generalization will prove to be central for some of our findings. Similarity to the existing literature in other dimensions is advantageous since it suggests that our findings do not arise due to non-standard assumptions. The second deviation from previous work is simply the addition of t subscripts to all parameters, reflecting our focus on the dynamic response to changes in the transmission framework. A parameter may change as of various underlying mechanisms. For example, an expansion of public childcare may affect the degree to which human capital is inherited across generations, or technological change may affect relative demand and thus returns to skills on the labor market. For simplicity we do not explicitly model any particular mechanism.

We will consider mobility trends following a single structural change in generation $t = T$, assuming that the moments of all variables were in steady-state equilibrium before the shock. For simplicity we assume that the process is infinite. This assumption (which imposes restrictions on the parameters of our model, see Appendix A.2) nor the existence of pre- and post-shock steady states are necessary for our arguments, but simplify the discussion and facilitate comparisons to the existing literature on steady-state mobility.

For convenience we normalize the variances of y_t and all elements of \mathbf{h}_t and \mathbf{e}_t in the initial steady state to one. The variances of $u_{y,t}$ and elements of $\mathbf{u}_{h,t}$ and \mathbf{v}_t are then implicitly a function of the slope parameters of the model, and the requirement for those variances to be non-negative leads to additional constraints on the parameters. Cross-sectional inequality may change after a structural change occurs. However, we will frequently consider changes in the *relative* strength of different transmission mechanisms that do not affect the cross-sectional variances of income, human capital, and endowments. Abstracting from changes in those variances simplifies the discussion and helps to isolate other adjustment mechanisms that are of particular interest.

2.2.1 The Importance of Past Transmission Mechanisms

We express intergenerational mobility as a function of our model to illustrate some central implications. Consider a simplified example, assuming $\mathbf{\Lambda}_t$ to be diagonal and cross-sectional inequality to remain constant, $Var(y_t) = Var(e_{j,t}) = 1 \forall j, t$. The intergenerational elasticity then coincides with the intergenerational correlation, and is derived by plugging equations (2.5) and (2.6) from our model into equation (2.1), such that

$$\beta_t = \frac{Cov(y_t, y_{t-1})}{Var(y_{t-1})} = \gamma_t + \boldsymbol{\rho}'_t \mathbf{\Lambda}_t Cov(\mathbf{e}_{t-1}, y_{t-1}). \quad (2.7)$$

Thus, β_t depends on current transmission mechanisms (parameters γ_t , $\boldsymbol{\rho}_t$ and $\mathbf{\Lambda}_t$) and on the cross-covariance between income and endowments in the parent generation. The intuition is simple. If income and other favorable endowments are concentrated in the *same* families then intergenerational mobility will be particularly low (the elasticity will be high). Expression (2.7) illustrates that two populations subject to the same transmission mechanisms (e.g., institutions and policies) today can still differ in their levels of intergenerational mobility, since current mobility depends also on the joint distribution of income and endowments in the parent generation.

The cross-covariance between income and endowments in the parent generation is in turn determined by past transmission mechanisms, and thus past values of $\{\gamma_t, \boldsymbol{\rho}_t, \mathbf{\Lambda}_t\}$. We can iterate equation (2.7) backwards to express β_t in terms of parameter values,

$$\begin{aligned} \beta_t &= \gamma_t + \boldsymbol{\rho}'_t \mathbf{\Lambda}_t (\mathbf{\Lambda}_{t-1} Cov(\mathbf{e}_{t-2}, y_{t-2}) \gamma_{t-1} + \boldsymbol{\rho}_{t-1}) \\ &= \dots \\ &= \gamma_t + \boldsymbol{\rho}'_t \mathbf{\Lambda}_t \boldsymbol{\rho}_{t-1} + \boldsymbol{\rho}'_t \mathbf{\Lambda}_t \left(\sum_{r=1}^{\infty} \left(\prod_{s=1}^r \gamma_{t-s} \mathbf{\Lambda}_{t-s} \right) \boldsymbol{\rho}_{t-r-1} \right), \end{aligned} \quad (2.8)$$

assuming that the process is infinite.¹¹ The level of intergenerational mobility today thus depends on current and past transmission mechanisms.¹² If no structural changes occur, $\gamma_s = \gamma$, $\boldsymbol{\rho}_s = \boldsymbol{\rho}$, $\mathbf{\Lambda}_s = \mathbf{\Lambda} \forall s \leq t$, then equation (2.9) simplifies to the *steady-state*

¹¹For a finite process, β_t will depend on past parameter values and the initial condition $Cov(\mathbf{e}_0, y_0)$.

¹²If cross-sectional inequality varies over generations, or if $\mathbf{\Lambda}_t$ is not diagonal, the derivation of equation (2.8) would require backward iteration of the variance of y_t and the variance-covariance matrix of \mathbf{e}_t . Accordingly, β_t would also depend on the variances of u_t and \mathbf{v}_t in past generations.

elasticity

$$\beta = \gamma + \boldsymbol{\rho}'\boldsymbol{\Lambda} \sum_{s=0}^{\infty} (\gamma\boldsymbol{\Lambda})^s \boldsymbol{\rho} = \gamma + \boldsymbol{\rho}'\boldsymbol{\Lambda} (\mathbf{I}_{KxK} - \gamma\boldsymbol{\Lambda})^{-1} \boldsymbol{\rho}, \quad (2.9)$$

where the second step follows since the geometric series $\sum_{s=0}^{\infty} (\gamma\boldsymbol{\Lambda})^s$ converges (the absolute value of each eigenvalue of $\gamma\boldsymbol{\Lambda}$ is below one). The literature has almost exclusively focused on how changes in structural parameters affect intergenerational mobility in steady state, as given by (2.9). We will instead analyze the *transition path* towards the new steady state as determined by equation (2.8).

Some properties can be readily generalized. The transition path of $Cov(\mathbf{e}_{t-1}, y_{t-1})$ is governed by the eigenvalues of the reduced-form coefficient matrix and is thus monotonic (see eq. 2.44 in Appendix A.2). But from (2.7) it follows that income mobility in the *first* generation subject to a structural change is directly affected by parameter changes, not indirectly by changes in the covariance between parental income and endowments. Trends in income mobility are thus not necessarily monotonic (even if cross-sectional inequality remains constant), as we will show in the next section. Other properties, such as the speed of convergence, depend on the parameterization of the model and can thus not be generalized.

2.2.2 From Simple Examples to Non-Monotonic Trends

We start with simplified versions of our baseline model and then move to more general models. For our first examples it is sufficient to consider a single endowment e_t and thus scalar versions of equations (2.5) and (2.6), such that

$$y_t = \gamma_t y_{t-1} + \rho_t e_t + u_t \quad (2.10)$$

$$e_t = \lambda_t e_{t-1} + v_t. \quad (2.11)$$

Our qualitative findings do not rely on specific parameter choices, but the quantitative implications of our examples will be more plausible if we choose values that are consistent with empirical evidence. The evidence in the literature, and our cross-validations within the model, suggest the following rough order of magnitudes for the US case:

$$0.45 \leq \beta \leq 0.55, \quad 0.15 \leq \gamma \leq 0.25, \quad 0.60 \leq \rho \leq 0.70, \quad 0.50 \leq \lambda \leq 0.65.$$

We discuss these choices in detail in Appendix A.3. It will be useful to first look at an even simpler case in which parental income has no causal effect.

Example 1: A simple meritocratic economy. Assume that the *heritability* of endowments (λ_t) or the *returns* to endowments and human capital (ρ_t) change in a simple meritocratic economy ($\gamma_t = 0 \forall t$).

Assume first that cross-sectional inequality remains constant.¹³ From equation (2.8), a change in the heritability of endowments in generation T from $\lambda_{t < T} = \lambda_1$ to $\lambda_{t \geq T} = \lambda_2$ shifts the intergenerational elasticity (or correlation) according to

$$\Delta\beta_T = \beta_T - \beta_{T-1} = \rho(\lambda_2 - \lambda_1)\rho. \quad (2.12)$$

Mobility remains constant afterwards. A change in returns from ρ_1 to ρ_2 in generation T instead shifts β_t over two generations. The first shift equals

$$\Delta\beta_T = \beta_T - \beta_{T-1} = (\rho_2 - \rho_1)\lambda \text{Cov}(e_{T-1}, y_{T-1}) = (\rho_2 - \rho_1)\lambda\rho_1, \quad (2.13)$$

and is induced by the change in returns for the *offspring* generation in T . The second shift,

$$\Delta\beta_{T+1} = \beta_{T+1} - \beta_T = \rho_2\lambda(\text{Cov}(e_T, y_T) - \text{Cov}(e_{T-1}, y_{T-1})) = \rho_2\lambda(\rho_2 - \rho_1), \quad (2.14)$$

is induced by the change in the correlation between income and endowments among the *parents* of the offspring generation $T + 1$, in turn caused by changing returns to those endowments in generation T . The second shift is larger than the first if returns increase ($\rho_2 > \rho_1$). Figure 1 gives a numerical example.

Cross-sectional inequality. An additional source of dynamics stems from changes in cross-sectional inequality. Intuitively, if *individual* endowments and skills are linked over generations due to inheritance within families, then *cross-sectional inequality* will also be linked over generations; the variance of equation (2.11) can be iterated backwards

¹³ Assume that the importance of parental background *relative* to unrelated factors changes, such that shifts in λ_t or ρ_t are offset by corresponding shifts in the variance of u_t or v_t .

such that

$$\text{Var}(e_t) = \lambda_{t-k}^{2k} \text{Var}(e_{t-k}) + \sum_{s=0}^{k-1} \lambda_{t-s}^{2s} \text{Var}(v_{t-s}) \quad \forall k \geq 1. \quad (2.15)$$

Models of intergenerational transmission therefore imply that the impact of a structural change on cross-sectional inequality may propagate in subsequent generations, in turn affecting mobility measures over multiple generations.¹⁴

Implications. The example illustrates that the dynamic response of mobility measures can be informative on the type of structural shock that occurred. Changes in the heritability of endowments and skills have a more immediate effect than changes in the returns to those skills, as income mobility depends directly on returns in both parent and offspring generations. The effect of changing returns on steady state mobility levels may thus not become fully evident before both the parent and child generations have experienced the new price regime. We can relate this argument to the evidence on rising skill differentials in wages from the late 1970s in the US, UK, and (more recently) other OECD countries. The notion that widening wage differentials could decrease intergenerational mobility (e.g., Blanden et al., 2004, and Solon, 2004) contributes greatly to the current interest in mobility trends. But recent studies do not yet observe offspring cohorts whose parents have fully experienced the changing wage regime; its impact on mobility may thus become more evident in future empirical work.¹⁵

Not only will the dynamic response of mobility depend on the type of structural change that occurred; different measures of the importance of family background may also show different dynamic responses. Sibling correlations, which capture influences on economic outcomes that are shared by siblings, depend less directly on conditions in the parent generation and thus respond more immediately to rising returns than intergenerational measures of persistence.¹⁶ This argument may explain why US studies find a sharp

¹⁴For example, if the changing heritability of endowments affects its cross-sectional variance (because the variance of v_t remains constant) then the elasticity shifts not only in the first but also subsequent generations, as

$$\Delta\beta_{T+1} = \rho\lambda_2 \left(\frac{\text{Var}(e_T)}{\text{Var}(y_T)} - \frac{\text{Var}(e_{T-1})}{\text{Var}(y_{T-1})} \right) = \rho\lambda_2 \left(\frac{1 + (\lambda_2^2 - \lambda_1^2)}{1 + \rho^2(\lambda_2^2 - \lambda_1^2)} - 1 \right)$$

is non-zero for $\lambda_1 \neq \lambda_2$.

¹⁵For example, the last offspring cohort observed in Lee and Solon (2009) were born in 1975. Their parents were not subject to the widening skill differential in their early careers.

¹⁶The sibling correlation equals $\rho_1^2\lambda^2$ before and $\rho_2^2\lambda^2$ in generations after returns change in the example.

increase in *sibling* correlations since 1980 (Levine and Mazumder, 2007), while there seems to be less evidence for such shift in *intergenerational* measures of persistence. The former are directly affected by changing wage differentials, but the latter also depend on conditions in the parent generation. Sibling correlations may then be a preferred measure in the analysis of mobility trends over time, as they tend to react more immediately to structural changes.¹⁷

These results have general implications for the interpretation of mobility trends: shifts in mobility may not reflect a changing effectiveness of current policies and institutions in the promotion of equality of opportunity, but a lagged effect of major changes in the more distant past. The next example illustrates that such repercussions can be both sizable and non-monotonic. We move to a more general model that allows for parental income to have causal effects ($\gamma \neq 0$). Consider first an example of “equalizing opportunities”, in which offspring outcomes become less dependent upon parental income.¹⁸

Example 2: Equalizing opportunities. Assume that the importance of parental status diminishes ($\gamma_1 > \gamma_2$) while skills that are partially inherited are instead more strongly rewarded ($\rho_1 < \rho_2$).

In other words, assume that in generation T the economy becomes less *plutocratic* and more *meritocratic*. For example, parental status may become less and own merits more important for appointment into jobs and occupations. Mobility then shifts in the first affected generation according to

$$\Delta\beta_T = (\gamma_2 - \gamma_1) + (\rho_2 - \rho_1)\lambda\text{Cov}(e_{T-1}, y_{T-1}), \quad (2.16)$$

affected both by the declining importance of parental income and the increasing returns to endowments or skills. However, the latter effect is attenuated, for two reasons. First, endowments are only imperfectly correlated within families, such that $\lambda < 1$. Second, parental endowments e_{T-1} explain only a fraction of the variation of incomes in the parent generation, such that $\text{Cov}(e_{T-1}, y_{T-1}) < 1$. Income mobility thus tends to increase if a generation is subject to a more meritocratic setting than their parents, as might be expected.

¹⁷Analysis of trends in sibling correlations, with its weaker data requirements, may also often be more feasible (see Björklund et al., 2009).

¹⁸As noted by Conlisk (1974a), “opportunity equalization” is an ambiguous term that may relate to different types of structural changes in models of intergenerational transmission.

However, income mobility will also shift in the second generation, according to

$$\Delta\beta_{T+1} = \rho_2\lambda \left[\frac{Cov(e_T, y_T)}{Var(y_T)} - \frac{Cov(e_{T-1}, y_{T-1})}{Var(y_{T-1})} \right]. \quad (2.17)$$

Apart from changes in the variance of income, the elasticity may also shift because of changes in the correlation between income and endowments in the parent generation. The relative importance of parameter changes on the latter is now reversed, since

$$\frac{\partial Cov(e_T, y_T)}{\partial \gamma_2} = \lambda Cov(e_{T-1}, y_{T-1}) \quad \text{and} \quad \frac{\partial Cov(e_T, y_T)}{\partial \rho_2} = 1.$$

Changing returns have a strong effect on the correlation between own endowments and incomes. A change towards a more meritocratic society tends to increase the correlation between endowments and income, thereby *decreasing* income mobility from the second affected generation onwards.

The dynamic response of the intergenerational elasticity thus tends to be *non-monotonic*, with an initial rise in mobility and a subsequent decline. Intuitively, a rise in the importance of own skill relative to parental status will be detrimental for offspring with high-income, low-skill parents. In contrast, the shift will benefit talented offspring from poor families, providing opportunities for upward mobility that were not yet available to their parents. Mobility is thus highest when these relative gains and losses occur, when a generation faces new institutions, policies and opportunities that differ markedly from those in their parents' generation. But the offspring of those who thrived under the meritocratic setting will also do relatively well, due to the inheritance of talent; mobility hence decreases subsequently.¹⁹

Exact conditions for such non-monotonic adjustment can be given if the shifting importance of parental background and own characteristics does not affect cross-sectional inequality, such that $Var(y_t) = 1 \forall t$.²⁰ Figure 2.2 plots a numerical example, illustrating that the response in mobility trends can be long-lasting; it becomes insignificant only in

¹⁹The idea that a shift towards “meritocratic” principles can also have depressing effects on mobility was already noted by the sociologist Michael Young, who coined the term in the book *The Rise of the Meritocracy* (1958). In contrast to its usage today, Young intended the term to have a derogatory connotation.

²⁰From equation (2.8), a change to a more meritocratic society will then increase mobility initially iff $\frac{\gamma_1 - \gamma_2}{\rho_2 - \rho_1} > \lambda Cov(e_{T-1}, y_{T-1})$. However, mobility *decreases* in subsequent generations iff $\frac{\rho_2 - \rho_1}{\gamma_1 - \gamma_2} > \lambda Cov(e_{T-1}, y_{T-1})$. These conditions will be satisfied for any changes $\gamma_1 - \gamma_2$ and $\rho_2 - \rho_1$ that are of similar magnitude in absolute terms.

the third generation, or more than half a century after the structural change.²¹

Implications. The example illustrates that we need to be careful when interpreting mobility trends. Not only may those trends be a response to events that occurred in past generations, this response may also be non-monotonic. Changes that are mobility-enhancing in the long run may nevertheless cause a *decreasing* trend in mobility measures that lasts over several generations. Declining mobility today may then not necessarily reflect a recent deterioration of equality of opportunity, but rather major gains made in the past.

In the numerical example, mobility responded much more strongly in the first two than in subsequent generations. Can we then conclude that more distant events have only a negligible effect on current trends? We believe not, for two reasons. First, plausible extensions of our model would generate slower transitions between steady states (e.g., considering wealth or capital accumulation, and direct causal effects from grandparents). Second, past events may have been more dramatic than more recent changes. For example, in the late 19th and early 20th century the US experienced rapid industrialization and urbanization, a strong decline in agricultural employment, mass migration, and a vast expansion of public schooling. The US participated in two world wars and went through a highly turbulent interwar period. Other countries experienced similarly stark transformations.

Much of the recent empirical literature measures trends in income mobility for offspring cohorts born from around 1950 to the 1970s, which are separated by only one or two generations from those events. Recent trends may thus partly reflect repercussions from such changes in the first half of the 20th century. Finally, our example illustrates that if those changes led to a more meritocratic society, mobility should perhaps be *expected* to decline in more recent cohorts.

2.2.3 Intergenerational Mobility in Times of Change

Our finding that a change to a more meritocratic society can lead to long-lasting and non-monotonic mobility trends is important for the interpretation of recent trends. But it relates to a rather specific structural change; one may thus expect that non-monotonic responses are more of an exception than a rule.

²¹We will illustrate the timing of mobility trends over *cohorts* further in Section 2.4.

We next illustrate that such responses are instead quite typical. We now consider multiple types of human capital and endowments, as in equations (2.5) and (2.6). The notion of individual ability has recently shifted from a one-dimensional concept primarily related to IQ (as in [Herrnstein and Murray, 1994](#)) to a multidimensional set of traits that also recognizes the importance of noncognitive skills. A stream of evidence has supported this idea, showing that several distinct skills affect various labor market outcomes (e.g., [Heckman et al., 2006](#); [Lindqvist and Vestman, 2011](#)). Such multiplicity has not yet been stressed in the intergenerational context (an exception is [Bowles and Gintis, 2002](#)), but our analysis illustrates that it provides implications that cannot be captured by single-skill models.²²

Example 3: Changing returns to skills. Assume that the returns to different types of human capital or endowments change on the labor market ($\rho_1 \neq \rho_2$).

Changes in the returns to different types of skills could stem from changes in demand (e.g., as of trade, or industrial and technological change) or in relative supplies (e.g., as of immigration or changes in the production of skills). A specific example is the decrease in the demand for physical relative to cognitive ability as a labor market moves from agricultural to white-collar jobs. But relative returns may change also in periods that are much shorter than the time scale underlying our intergenerational analysis – a typical example is the job-polarization literature, which highlights how the IT revolution has implied a shift in demand from substitutable manual skills to complementary abstract skills (e.g., [Levy, Murnane, and Autor, 2003](#)).

Figure (2.3) illustrates a simple symmetric case: two endowments k and l are equally transmitted within families ($\lambda_{ij} = \lambda$ for $i = j$ and $\lambda_{ij} = 0$ for $i \neq j$), but their prices on the labor market swap at time T ($p_{2,k} = \rho_{1,l} \neq p_{1,k} = \rho_{2,l}$). Adapting equations (2.5) and (2.6) for $K = 2$ endowments and iterating backwards we find that mobility increases in the first affected generation, but decreases in the next.²³

Intuitively, those endowments or skills that have been more strongly rewarded in past generations are also more strongly correlated with parental income. As a consequence,

²²Multiplicity of skills matters also for other questions in the literature. For example, in chapter 4 I show that income persistence over generations may decline more slowly than at a geometric rate if the degree of heritability varies across characteristics.

²³We find $\Delta\beta_T = -(\rho_{k,2} - \rho_{k,1})^2 \lambda / (1 - \gamma\lambda)$, which is negative. The elasticity in the second generation shifts according to $\Delta\beta_{T+1} = \lambda(\rho_{k,2} - \rho_{k,1})^2 + \lambda(\rho_{k,2}^2 + \rho_{k,1}^2 + (2\rho_{k,1}\rho_{k,2}\lambda\gamma)/(1 - \gamma\lambda))(1/\text{Var}(y_T) - 1)$, which is positive since $\text{Var}(y_T) = 1 - 2\gamma\lambda(\rho_{k,2} - \rho_{k,1})^2/(1 - \gamma\lambda) < 1$. These findings are not due to shifts in cross-sectional inequality; if instead $\text{Var}(y_T) = 1$ (i.e. changes in ρ_k and ρ_l are offset by changes in the variance of u_t) we still have that $\Delta\beta_T < 0$ and $\Delta\beta_{T+1} > 0$.

mobility tends to initially increase if relative prices change, since endowments for which prices increase from low levels are less prevalent among high-income parents than endowments for which prices decrease from high levels. But the endowment for which prices increase becomes increasingly associated with high parental income in subsequent generations, causing a decreasing mobility trend. The key assumptions underlying these results are that endowments are positively correlated within families and imperfectly correlated within individuals.

We can derive that non-monotonic responses in mobility are also typical when the returns to any number of skills change, by expressing the elasticity in generation T as a function of the steady-state elasticities before and after the structural change (β_{T-1} and $\beta_{t \rightarrow \infty}$). We assume here a diagonal heritability matrix $\mathbf{\Lambda}$. The derivation for more general cases (non-diagonal $\mathbf{\Lambda}$ and correlated endowments) is given in Appendix A.4. If the steady-state variance of income remains unchanged we have

$$\beta_{T-1} = \gamma + \boldsymbol{\rho}'_1 \mathbf{\Lambda} (\mathbf{I} - \gamma \mathbf{\Lambda})^{-1} \boldsymbol{\rho}_1 \quad (2.18)$$

and

$$\beta_{t \rightarrow \infty} = \gamma + \boldsymbol{\rho}'_2 \mathbf{\Lambda} (\mathbf{I} - \gamma \mathbf{\Lambda})^{-1} \boldsymbol{\rho}_2, \quad (2.19)$$

such that

$$\beta_T = \frac{1}{2} (\beta_{T-1} + \beta_{t \rightarrow \infty}) - \frac{1}{2} (\boldsymbol{\rho}'_2 - \boldsymbol{\rho}'_1) \mathbf{\Lambda} (\mathbf{I} - \gamma \mathbf{\Lambda})^{-1} (\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1). \quad (2.20)$$

The quadratic form in the last term is greater than zero for $\boldsymbol{\rho}_2 \neq \boldsymbol{\rho}_1$ since $\mathbf{\Lambda} (\mathbf{I} - \gamma \mathbf{\Lambda})^{-1}$ is positive definite. Eq. (2.20) states that intergenerational mobility in the first affected generation can be decomposed into two parts. Mobility in generation T equals the average of the old and the new steady-state mobility (first term), plus a purely *transitional gain* (second term). Price changes then lead to a temporary spike in mobility (β_T is below both the previous steady state β_{T-1} and the new steady state $\beta_{t \rightarrow \infty}$) if the steady-state elasticity does not shift too strongly, iff

$$|\beta_{t \rightarrow \infty} - \beta_{T-1}| < (\boldsymbol{\rho}'_2 - \boldsymbol{\rho}'_1) \mathbf{\Lambda} (\mathbf{I} - \gamma \mathbf{\Lambda})^{-1} (\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1). \quad (2.21)$$

This argument also holds if cross-sectional inequality is lower in the new than in the

old steady state.²⁴ Any symmetric changes (as in the numerical example) or changes in returns that do not affect long-run mobility much fulfill condition (2.21) and will thus lead to non-monotonic trends as in Figure 3.

We should thus expect “short-term” mobility gains if returns change, but those gains may not persist. These results have general implications on how we expect institutional or technological change to affect mobility. Previous authors have shown that technological progress can lead to non-monotonic mobility trends through *repeated* changes in skill returns (Galor and Tsiddon, 1997). We find that even a *one-time* change tends to generate such trends.

Implications. We can formulate a more general intuition, which applies to both of our last two examples. A change in the relative importance of different channels of intergenerational transmission will tend to increase mobility temporarily, as it affects the prospects of families differently. For example, a decline in the importance of parental income relative to own skills diminishes the prospects of offspring from high-income parents. The declining relative importance of a particular skill or endowment is to the disadvantage of those families in which it is abundant. Technological, economic, and social changes will often generate such relative gains and losses, generating *transitional* intergenerational mobility in the generation in which they occur.

The implications of our findings are not restricted to those particular types of structural changes that we examined explicitly. This may become more apparent if we allow for a broader definition of the endowment vector. For example, assume that \mathbf{e}_t captures also the geographic location of individuals (“inherited” with some probability from their parents). We can then relate our last example to Long and Ferrie (2013b), who argue that US occupational mobility may have been comparatively high in the 19th century as of exceptional internal *geographic* mobility. Our framework can support this hypothesis, but with a different emphasis. Intergenerational mobility may not necessarily increase due to internal migration itself (that depends on who migrates), but certainly due to one of its underlying causes: variation in labor demand across areas and time incentivizes internal migration, but it also directly increases intergenerational income mobility by generating different local demand conditions for parents and their (non-migrating) children.

²⁴Eq. (2.20) includes then the additional term $\boldsymbol{\rho}'_2 \boldsymbol{\Lambda} (\mathbf{I} - \gamma \boldsymbol{\Lambda})^{-1} \boldsymbol{\rho}_2 (1 - \frac{1}{\text{Var}(y_{t \rightarrow \infty})})$, which is negative if $\text{Var}(y_{t \rightarrow \infty}) < \text{Var}(y_{T-1}) = 1$.

We thus come to a quite general conclusion. First, times of change tend to be times of high intergenerational mobility. Moreover, such gains will be succeeded by a long-lasting decline in mobility, unless further structural changes occur. Countries experiencing a period of stable economic conditions will thus tend to be characterized by negative mobility trends if they were preceded by more turbulent times.

As noted above, countries such as the US may have experienced much greater societal transformations in the first than in the second half of the 20th century. Our findings suggest that such transformations may have strengthened intergenerational mobility in economic status in those generations that were directly affected.²⁵ Our model also illustrates that these mobility gains diminish in subsequent generations, providing another reason why mobility of more recent cohorts should perhaps be *expected* to decline.

2.3 Empirical Application

The core implication from our model is that even a single structural change should be expected to affect intergenerational mobility measures over long time periods. We examine now if such dynamic effects can be observed empirically.

We considered intergenerational mobility trends over *generations* in our theoretical framework, but empirical studies estimate mobility trends over *cohorts* (typically offspring cohorts). These two dimensions, which do not match due to variation of parental age at birth, have to our knowledge not yet been linked in the literature. An explicit consideration of cohorts (Section 2.4) will provide additional implications, some of which will already become apparent in our empirical analysis.

Our objective is to cleanly identify the effects of a major structural reform on mobility not only in the directly affected cohorts, but also in subsequent cohorts and generations. This intention leads to considerable requirements on both data coverage (requiring data on family links and individual outcomes over multiple decades) and identifiability of the reform impact among other determinants of mobility trends. Fortunately, the Swedish compulsory school reform and access to long-run registry data make such analysis possible.

²⁵Note that much of the economic literature and our findings relate to *relative* mobility, how differences in economic outcomes among parents relate to differences among their offspring. Economic development or transitions may also generate *absolute* mobility, by generating differences in economic status between generations (see Goldthorpe, 2013).

2.3.1 The Swedish Compulsory School Reform

We describe here only the most important elements of the Swedish compulsory school reform, which is comprehensively discussed in [Holmlund \(2007\)](#). Gradually implemented across municipalities from the late 1940s, the reform’s two main components were to *raise compulsory schooling* from seven (eight in some municipalities) to nine years, and to *postpone tracking* decisions from the fifth or seventh to after the ninth grade. The reform prescribed a unified national curriculum and municipalities received additional funding to cover costs from its implementation.

Our choice of application is motivated by three main reasons. First, education and educational systems are key mechanisms for the reproduction of economic advantage. Family background explains a large share of the variation in educational attainment, and institutional aspects are believed to affect that relationship ([Björklund and Salvanes, 2010](#)). Educational reforms or expansion are thus potential determinants of observed mobility changes over time ([Machin, 2007](#)), and school reforms are often directly motivated by a desire to increase mobility – indeed, one of the Swedish reform’s objectives was to increase educational attainment among students from less advantaged backgrounds ([Erikson and Jonsson, 1996](#)). The Swedish and similar reforms in other Scandinavian countries have appeared to achieve this objective, raising income mobility in directly affected generations (see [Meghir and Palme, 2005](#), [Holmlund, 2008](#), and [Pekkarinen et al., 2009](#)).

Second, administrative data in Sweden cover an extraordinarily long time span. Coverage over three generations is needed to assess the reform’s impact on mobility not only on directly affected but also the subsequent generation. Large sample sizes allow us to exploit fine geographic variation for causal identification and to detect gradual mobility changes over time.

Third, the reform’s gradual implementation over municipalities allows separation of the reform from regional or time-specific effects. A number of studies exploit this characteristic to assess the causal effect of the reform on individual outcomes in directly affected, or spillover effects in subsequent generations (see e.g. [Meghir and Palme, 2005](#); [Holmlund et al., 2011](#); [Meghir et al., 2011](#)). While we follow a similar identification strategy, our objective is to examine the reform’s effect on standard summary measures of intergenerational mobility instead of individual outcomes. Both aspects are related

(e.g., [Havnes and Mogstad, 2012](#)), but mobility can respond dynamically even in the absence of intergenerational spillover effects, as we showed theoretically in Section 2.2.2.

We estimate the reform’s impact on intergenerational mobility in income and educational attainment over two generations and compare the results against our theoretical predictions.

2.3.2 Compulsory Schooling in the Intergenerational Model

The impact of a compulsory schooling policy on educational and income mobility can be predicted from our theoretical framework. We first include constants α_y and α_h into the scalar variants of our baseline equations (2.2)-(2.3), thus allowing for mean changes in income and education. To capture the main component of the school reform assume then that eq. (2.3) determines *intended* schooling h^* , while from generation T onwards *actual* schooling h_t is compulsory until x years, such that

$$h_t = \begin{cases} h_t^* & \text{if } t < T \\ \max(h_t^*, x) & \text{if } t \geq T \end{cases}. \quad (2.22)$$

The school reform raises schooling of individuals with particularly low educational attainment. This “mechanical” shift may in turn affect the attainment of others via potential general equilibrium responses. Compositional changes may generate peer effects, and changes in supply may alter the returns to schooling and thus schooling decisions.²⁶ However, a theoretical discussion of the numerous responses that may occur over such long time intervals can be only incomplete and speculative. We instead focus on the main “mechanical” effect of the school reform, which explains the observed empirical pattern well.

We study the dynamic response in the most popular measure of income and educational mobility, the intergenerational elasticity of income β_{inc} and educational coefficient β_{edu} ,

$$\beta_{inc,t} = \frac{Cov(y_t, y_{t-1})}{Var(y_{t-1})} \quad \text{and} \quad \beta_{edu,t} = \frac{Cov(h_t, h_{t-1})}{Var(h_{t-1})}. \quad (2.23)$$

In the previous section we derived this measure by repeated insertion of the structural

²⁶Spillover effects on educational attainment of individuals not directly affected by the reform were found to be small in [Holmlund \(2007\)](#).

equations of our model, using linearity of the expectation operator to solve for the required moments. But the compulsory schooling requirement generates a non-linear relationship between h_t and h_{t-1} , which depend also on the distributions of u_y , u_h and v .

Figure 2.4 provides a simulated numerical example based on simple parametric assumptions (e.g., normally distributed errors). From generation T schooling becomes compulsory until $x = 9$ years. We assume that parental schooling has only modest indirect intergenerational spillover effects ($\gamma_h = 1$) and choose other parameters such to generate pre-reform first and second moments for income y_t and schooling h_t that are similar to the observed moments in the Swedish data.

Panel A plots the response of the intergenerational educational coefficient β_{edu} . In offspring generation T the reform compresses the variance of schooling strongly, which decreases the numerator of β_{edu} – differences in schooling between parents result into smaller differences among their offspring. However, from generation $T + 1$ the variance of schooling is also compressed among *parents*, who were already subject to the school reform in the previous generation. The coefficient β_{edu} is inversely scaled by this variance, and thus tends to rise. The non-monotonic response is thus mainly a consequence of strong changes in the variance of the marginal distributions (a direct and mechanical effect of the reform).

The reform could lead to further substantial compressions of educational attainment in subsequent generations if schooling has very strong causal effects on offspring outcomes ($\gamma_h \gg 1$). However, the existing empirical literature points to modest intergenerational “multiplier” effects of education (see Plug et al., 2011). The dashed line illustrates one important potential general equilibrium response. Increased supply of formal schooling may decrease its returns on the labor market (a decrease in δ), decreasing inequality in income and thus (if human capital accumulation is subject to parental investments) educational inequality and intergenerational persistence.

A reduction in the degree to which differences in educational attainment are transmitted from parents to offspring will also reduce the transmission of income differences, if formal schooling improves an individual’s earnings potential – the intergenerational income elasticity β_{inc} decreases in generation T (panel B in Figure 2.4). General equilibrium responses may affect this prediction. For example, increased supply of formal schooling may reduce its returns, thus decreasing the intergenerational elasticity further (dashed line). The second-generation response in β_{inc} is less clear-cut. Changes in the numerator

of β_{inc} in eq. (2.23) are not as easily dominated by a decrease in the denominator in generation $T + 1$, which will tend to be weaker for β_{inc} than for β_{edu} since differences in formal schooling are not the only source of differences in income. The direction of the second-generation response in β_{inc} is thus an empirical question.

2.3.3 Data

Our *source data* is based on a 35 percent random sample of the Swedish population born between 1932 and 1967. Using information based on population registers we link sampled individuals to their siblings (all sibling types) as well as their (and their siblings') biological parents and children. We then individually match data on personal characteristics and place of residence based on bi-decennial censuses starting from 1960, as well as education data stemming from official registers. We do not use the sibling-parent subsample in our main analysis: it can provide additional precision in mobility estimates in 1940/50 cohorts, but is not representative for earlier and later cohorts.

Educational registers were compiled in 1970, 1990 and about every third year thereafter, containing detailed information on each individual's educational attainment.²⁷ Data in 1970 were collected only for those born 1911 and later. We can therefore not observe schooling for parents who were 33 years or older at their child's birth in 1943 (at the onset of the reform implementation). This age limit increases by a year for each subsequent offspring cohort, potentially creating a confounding trend in mobility measures over cohorts due to non-random sample selection. For comparability we thus restrict our *intergenerational sample* to parent-child pairs in which parents were no older than 32 years when their child was born. Educational data may also be missing for other reasons, in particular if parents had died or emigrated before 1970. The probability of such occurrences is potentially related to individual characteristics, but the share of affected observations is small.²⁸ As the data are collected from official registers there are no standard non-response problems.

The most recent educational register was compiled in 2007, which allows us to consider

²⁷We consider for each individual the highest attainment recorded across these years. The information on schooling levels is translated into years of education with 7 years for the old compulsory school being the minimum, and 20 years for a doctoral degree the maximum.

²⁸Educational information are less often missing among offspring, due to their younger age and the more frequent measurement of education after 1990. The share of missing observations does not vary with reform status (conditional on municipalities and offspring cohorts), and has thus little effect on our causal analysis.

mobility trends for cohorts born from the early 1940s up until 1972. Attainment of individuals at the top of the educational distribution is not reliably covered for more recent cohorts; only a small population share is affected, but measurement error in the tails of the distribution would have a disproportionately large effect on intergenerational mobility measures.

We construct a measure of long-run income status based on age-specific averages of annual incomes, which are observed for the years 1968-2007.²⁹ Incomes for parents are necessarily measured at a later age than incomes for their offspring, which may bias estimates of the intergenerational elasticity of lifetime income. Such bias is less problematic for our purposes as we are interested in mobility differences between groups instead of the overall level of income mobility in the population.

We present evidence on mobility in father-child pairs, but the consideration of maximum parental education and income yields similar results. We test the robustness of our results using other samples with no or different restrictions on parental age, or alternative measures of parental education and income, some of which we will also report below.

To construct the reform dummy, which indicates whether an individual was subject to the new system of comprehensive schooling, we follow the procedure first used by [Holmlund \(2008\)](#). Reform status can be approximated using information on an individual's birth year (from the administrative register) and place of residence during school age (from the censuses).³⁰ The gradual implementation of the reform affected cohorts born between 1938 and 1955, but the school municipality cannot be reliably determined for individuals born before 1943. As the share of individuals affected by the reform was very small we set the reform dummy to zero for all cohorts before 1943 (and one for all cohorts after 1955).³¹

Table 2.1 describes, by birth cohort, both the source data and the intergenerational sample, which was drawn according to the conditions described above. The number of

²⁹We use total (pre-tax) income, which is the sum of an individual's labor (and labor-related) earnings, early-age pensions, and net income from business and capital realizations. We express all incomes in 2005 prices and exclude observations with average incomes below 10000 SEK.

³⁰Reform status across cohort-municipality cells can be inferred by tracing in which cohort, for each municipality, the share graduating from the old school system discontinuously drops to zero (or close to zero). Helena Holmlund has kindly provided us with her coding, and we refer to [Holmlund \(2007\)](#) for further details on the coding procedure and potential measurement issues.

³¹Cohorts born before 1943 were subject to the new school system in 33 out of a total of 1034 municipalities. With the exception of less than a handful mid-sized urban municipalities, all of these were small, rural municipalities. We further drop a small number of municipalities for which the implementation date is unclear.

observations for each cohort are listed in columns 2 and 5. Columns 6 and 7 describe the number of observations with non-missing education or income information. Columns 3-4 and 8-9 describe how the share of offspring and fathers attending reformed schools increases over cohorts. It increases faster among fathers in the intergenerational sample than in the source data, due to oversampling of younger parents in the former.³²

2.3.4 Empirical Evidence

Descriptive Evidence. To illustrate the timing of the reform further, Figure 2.5 plots the shares of offspring and fathers attending a reformed school in our source data over half a century of (offspring) birth cohorts. The share of children subject to the reform increases nearly linearly in cohorts 1943-1955 (gray area). These individuals become parents themselves from the early 1960s, but their share among all parents increases more slowly due to variation in parental age at birth (black area). Up until the early 1980s only a minority of fathers had themselves been affected by the compulsory school reform. This observation leads to a first important point: the dynamic effect of structural changes on mobility measures in subsequent generations should be *gradual*, due to variation of parental age at birth. We will discuss this implication in more detail in Section 2.4. As noted, the share of fathers subject to the reform increases faster in our intergenerational sample, which is restricted to younger parents (dashed line). Our results will therefore understate the longevity of the reform’s effect on mobility measures. The reform had a direct impact on educational attainment, which can be also measured with high precision over long time intervals.³³

Figure 2.6 plots the mean and variance of years of schooling of offspring cohorts (1933-1972) and their fathers (1911-1935) in our intergenerational sample. Vertical bars at the 1943 and 1955 cohorts indicate the start and end point of the reform’s implementation. A reform effect on *average* years of schooling is not easily discernible from panel (A). Indeed, [Holmlund \(2007\)](#) finds the reform effect on mean schooling to be small (lower

³²A smaller share of individuals from the raw data are sampled among earlier cohorts, as their fathers are less likely to be identified in the source data. Identification of the reform effect requires that the probabilities that fathers, education and income are observed do not change systematically with introduction of the reform. While sampling probabilities differ across birth cohorts and municipalities, the correlation with reform status is negligible.

³³A measure of education in later life is likely to capture an individual’s entire educational attainment, as most people complete schooling in early life. In contrast, differences in current incomes are poor proxies of differences in lifetime income, such that measures of income mobility (in particular of mobility trends) are sensitive even to small changes in the age at which incomes are observed (the *life-cycle bias* problem, see [Jenkins, 1987](#), [Haider and Solon, 2006](#), and [Nyblom and Stuhler, 2011](#)).

bound estimate of 0.19 years), as only a share of children are affected by the compulsory requirement. In contrast, the shift in the *variance* of schooling is more striking: the reform period coincides with a sudden and strong compression of the distribution of schooling. Comparison with earlier trends for their fathers in the first half of the 20th century illustrates the exceptional magnitude of those changes.

Intergenerational Mobility Trend. Figure 2.7 plots cohort trends in the intergenerational educational coefficient, the slope coefficient in an ordinary least-squares regression of offspring’s years on father’s years of schooling. The solid line includes estimates from our main intergenerational sample, spanning from 1943 to 1972. The dashed line represents estimates from a restricted sample containing younger fathers (aged below 30), allowing us to plot trends also for earlier cohorts not yet affected by the reform. We find estimated trends to be very robust to changes in sample restrictions concerning parental age, as exemplified by the close overlap for the 1943-1945 cohorts (plotted) and beyond.

The reform’s implementation period coincides with a large drop in the intergenerational coefficient, contrasting with stable estimates before the onset of the school reform. The degree to which differences in schooling are transmitted to the next generation declines by more than a third. This decline is consistent with our theoretical expectation: the reform compresses the distribution of years of schooling in the offspring generation, such that differences in parental education correspond to smaller differences in offspring attainment.

Reform Effect. Figure 2.8 provides more direct evidence on the reform impact. Recentering the data within each municipality, we compare educational attainment and the intergenerational educational coefficient before and after a cohort was first subject to the new school type. The share of individuals with less than 9 years, the variance of schooling and the intergenerational schooling coefficient all drop strongly with local reform implementation.

We can exploit the gradual introduction of the reform to verify its causal impact, adapting a difference-in-differences specification as similarly used in [Holmlund \(2008\)](#)

and Pekkarinen et al. (2009). Consider the regression equation for schooling (income)

$$\begin{aligned}
 h_{cfm,t} = & \underbrace{\alpha_1 + \beta_1 h_{t-1}}_{\text{baseline}} + \underbrace{\alpha_2 R_{cm} + \beta_2 (h_{t-1} \times R_{cm})}_{\text{reform effect}} \\
 & + \underbrace{\alpha'_3 \mathbf{D}_c + \beta'_3 (h_{t-1} \times \mathbf{D}_c)}_{\text{offspring cohort effect}} + \underbrace{\alpha'_4 \mathbf{D}_m + \beta'_4 (h_{t-1} \times \mathbf{D}_m)}_{\text{municipality effect}} + \varepsilon_{cfm,t}, \quad (2.24)
 \end{aligned}$$

where $h_{cfm,t}$ represents years of schooling (log income) of the offspring in generation t of family i (subscript suppressed) born in cohort c , to a father of generation $t - 1$ born in cohort f , attending school in municipality m . The variable h_{t-1} represents years of schooling (log income) of fathers. The indicator R_{cm} equals one if the reform was in effect for cohort c in municipality m . We control for differences in both schooling levels and the intergenerational coefficient across child cohorts (captured in the indicator vector \mathbf{D}_c) and across municipalities (captured in \mathbf{D}_m).

The identifying variation that we exploit in this specification are municipality-specific changes in the intergenerational coefficient after local introduction of the reform. While controlling for common time trends and for persistent differences across areas, this strategy is still susceptible to differences in municipality-specific trends. Moreover, the reform indicator is measured with error, which may introduce attenuation bias. We address both issues below.³⁴

Table 2.2 presents OLS estimates from different variants of specification (2.24), based on a pooled sample of those cohorts that were affected by the reform introduction phase (1943-1955). Panel (A) presents our findings on educational mobility. The estimated schooling coefficient for a simple pooled regression (column 1) of 0.359 approximates the average of cohort-specific estimates over that period (see Figure 2.7).³⁵ The second column presents separate estimates for children who were and who were not subject to the reform. Differences in parental educational attainment are associated with much smaller differences in attainment among the former. To identify the reform's causal contribution we successively introduce cohort and municipality fixed effects and interactions in the next columns. Standard errors are clustered on the municipality level. Estimates for the full difference-in-differences specification are presented in column 4. We find that the

³⁴Some pupils may have moved in response to local reform implementation, but Holmlund (2007) finds that there was little selective mobility with respect to parental background.

³⁵Differences in yearly means also affect the pooled coefficient (Hertz, 2008), but their contribution is small.

Swedish compulsory school reform reduced the degree to which differences in educational attainment were transmitted from fathers to their children by about ten percent ($\hat{\beta}_2 = -0.0371$, $p < 0.001$).

Panel (B) of Table 2.2 presents corresponding estimates on income mobility. Our measure of long-run income of offspring (fathers) is based on average incomes in age 30-35 (age 53-59). Given observation of incomes at such a young (old) age for offspring (fathers), the pooled coefficient of 0.164 is likely to understate the true degree of intergenerational persistence in lifetime income (see chapter 3). We can nevertheless identify if the reform had an effect on income mobility. Our difference-in-differences estimate implies that the degree to which percentage income differences were transmitted from fathers to their children decreased by about ten percent due to the reform ($\hat{\beta}_2 = -0.0196$, $p < 0.05$). These results are consistent with findings by [Holmlund \(2008\)](#).

Our estimates are not sensitive to the inclusion of father cohort effects and remain statistically significant also for a number of alternative specifications, as discussed in more detail below. We conclude that the reform had a clear positive effect on both educational and income mobility in the first affected generation.

Heterogeneity. Yet, this effect may be smaller than expected. The intergenerational educational coefficient dropped by more than a third during the reform introduction phase (from about 0.42 to 0.27, see Figure 2.7). Furthermore, a sudden trend change occurred in the mid-1940s, even though few municipalities had yet been subject to the reform. This pattern can be understood if we examine the heterogeneity in the reform's effect over time. We interact the reform with offspring cohort dummies, exploiting that in each cohort additional municipalities switch to the new school system. The reform effect in specification (2.24) then equals

$$\alpha_2 (R_{cm} \times D_c) + \beta_2 (h_{t-1} \times R_{cm} \times D_c). \quad (2.25)$$

Figure 2.9 plots the resulting estimates for the elements of β_2 (black line). The reform had a very strong impact in earlier cohorts, reducing the intergenerational coefficient by almost 25 percent in those municipalities that were subject to the reform already in the early 1940s. But coefficient estimates decrease over cohorts, implying that its impact on later cohorts was small. The reason becomes clear from Figure 2.6. The general trend

towards higher educational attainment made the main component of the reform (the rise of the compulsory school level to nine years) less consequential: as by the early 1950s most pupils were attending school for at least nine years anyways. The reform effect can thus be seen as an intention-to-treat estimate, with the share of compliers diminishing over cohorts. We therefore conclude that the reform caused the sudden drop in the intergenerational coefficient in the early 1940s, but that much of its overall decline until the mid 1950s might have occurred even in the absence of the reform.³⁶

The pooled (difference-in-differences) coefficient that we presented in Table 2.2 can be decomposed as a weighted average of these cohort-specific reform effects,

$$\beta_2(DD) = \sum_{c=1}^{\bar{c}} \beta_{2,c} w(\beta_{2,c}), \quad (2.26)$$

where c denotes cohorts, $\beta_{2,c}$ the cohort-specific reform effects, and $w(\beta_{2,c})$ the weight assigned to each cohort. These weights are defined as

$$w(\beta_{2,c}) = \frac{Var(h_{t-1} \times R_{cm} \mid h_{t-1}, R_{cm}, \mathbf{D}_c = D_c) P(\mathbf{D}_c = D_c)}{\sum_{c=1}^{\bar{c}} Var(h_{t-1} \times R_{cm} \mid h_{t-1}, R_{cm}, \mathbf{D}_c = D_c) P(\mathbf{D}_c = D_c)}. \quad (2.27)$$

The pooled estimator will assign more weight to large cohorts, and cohorts with greater variance in father's schooling and the reform dummy (conditional on their covariance). Thus, the pooled coefficient is likely to be most affected by cohorts in which the shares of affected and unaffected by the reform are similar in size (i.e. the variance of the reform dummy is maximized). This will especially hold if the variance of father's schooling is relatively stable over the implementation period, which is true in our case. Sample analogs of these weights are plotted in the grey line in Figure 2.9. As suspected, the weights are highest around the 1950 cohort and still high for later cohorts. In contrast, the weights are close to zero for earlier cohorts. The pooled coefficient (Table 2.2) reflects therefore mostly the reform impact on later cohorts, which was comparatively small.

Our example points to a general feature of difference-in-differences analyses with gradual (or staggered) treatment implementation. Treatment effects are assumed to be constant over time in a standard specification, but are likely heterogeneous if the

³⁶Our estimates may understate the reform effect in later cohorts if it generated anticipation or spillover effects in non-reform schools (individual schooling decisions may depend on the educational attainment of others). Our argument that the reform's impact was larger in earlier cohorts still holds, as educational attainment was steadily increasing even before the reform was introduced.

counterfactual is subject to trends. The pooled coefficient then gets a more complex interpretation and may to a large extent reflect the reform impact at a particular point in time, which can be quite different from its initial impact.

Second generation effect. Figure 2.7 documents a second, more gradual but nevertheless pronounced change in mobility over time. After its large and long decline, the coefficient starts rising again among cohorts born in the late 1960s. Incidentally, these were the first cohorts in which some children were born to fathers who already themselves attended a reform school (see Figure 2.5).

But is the modest increase in the data really the dynamic impact of the reform, and not the product of coincidental (and potentially contemporaneous) factors? We can distinguish these sources by adapting regression equation 2.24 for the next generation.

We estimate

$$\begin{aligned}
 h_{cfm',t} = & \underbrace{\alpha_1 + \beta_1 h_{t-1}}_{\text{baseline}} + \underbrace{\alpha_2 R_{fm'} + \beta_2 (h_{t-1} \times R_{fm'})}_{\text{father reform effect}} \\
 & + \underbrace{\alpha'_3 D_f + \beta'_3 (h_{t-1} \times D_f)}_{\text{father cohort effect}} + \underbrace{\alpha'_4 D_{m'} + \beta'_4 (h_{t-1} \times D_{m'})}_{\text{father municipality effect}} + \varepsilon_{cfm,t}, \quad (2.28)
 \end{aligned}$$

where the indicator $R_{fm'}$ equals one if the reform was in effect for *father* cohort f born in municipality m' .

Table 2.3 presents OLS estimates from variants of specification (2.24), using offspring cohorts 1966-1972 in which the share of reform fathers is above one percent (adding earlier cohorts has little effect on the estimates). Panel (A) presents our results on educational mobility. Estimates for the full difference-in-differences specification (column 4, $\hat{\beta}_2 = 0.0655$, $p < 0.001$) indicate that the observed rise in the intergenerational educational coefficient is indeed a dynamic response to the school reform that occurred in the previous generation.

Panel (B) presents estimates of the reform's second-generation effect on income mobility. We can observe parental incomes at an earlier age for later cohorts, and use observations in age 35-45 to construct our measure of long-run status. The pooled coefficient estimate of 0.207 is thus likely to be less biased than the corresponding estimate for the first generation. As with education, the reform's impact on the intergenerational coefficient ($\hat{\beta}_2 = 0.041$, $p < 0.05$) is larger than the corresponding estimate for the first

generation. Two factors explain this finding. First, fathers who themselves were subject to the reform had their children at young age. Young fathers tend to have less educational attainment, the reform impact on this group was thus large. A second explanation follows from Figure 2.9 – children born in the late 1960s are more likely to have parents born in the early 1940s than later. We showed that the reform impact was much larger on the former, due to underlying trends in education.

In our data we can track the intergenerational coefficient only up to 1972, but the share of reform fathers will continue to climb until the early 2000s (see Figure 2.5). Unless dominated by contemporaneous events we thus expect the intergenerational income elasticity and educational coefficient to rise for several decades after our records end.

Intergenerational Correlation. The reform’s impact on the intergenerational regression coefficient exemplifies our argument that current mobility levels and trends can be affected by events that occurred in a more distant past. The school reform compresses the distribution of years of schooling, first decreasing the regression coefficient when affecting the offspring’s distribution, and increasing it in later cohorts when also affecting parents. But it is less obvious what trend we should expect in the intergenerational correlation, which abstracts from differences in cross-sectional inequality over generations.

Figure 2.10 plots estimates of the intergenerational correlation from 1940 to 1972. Estimates from our main intergenerational sample are represented by the solid line, while the dashed line shows estimates from a restricted sample containing fathers aged below 30 to examine trends also for earlier cohorts. Estimated levels are sensitive to changes in sample restrictions concerning parental age, but the pattern over cohorts appears robust. The intergenerational correlation is strongly increasing among cohorts not yet affected by the reform, but the correlation starts declining shortly after introduction of the reform from 1943 and remains lower until the end of our observation period in the early 1970s. The overall change in the correlation is smaller than the change in the regression coefficient.

Estimates are comparatively low already in the early 1950s. The difference is not statistically significant, but such pattern would not be surprising: our model predicts that the intergenerational correlation should be particularly low when the shares of children subject and not subject to the reform are similar, as a larger part of the variation in

schooling is then explained by reform status instead of parental background. The rising coefficient towards the end of our sampling period is not predicted by our model; given its suddenness it is likely due to contemporaneous instead of past events. We return to this argument in our next section.

Robustness. We perform a number of tests to probe the robustness of our results. Table 2.4 compares our baseline estimates of the reform effect on the intergenerational educational coefficient and income elasticity with estimates from six alternative specifications. First, we include matched siblings in our sample, which increases its size but also diminishes representativeness for some cohorts (see data subsection). Second, we restrict the sample to younger fathers with age at birth below 30, to probe the sensitivity of our results to such age restrictions. Our third robustness tests address measurement error in the reform indicator. Individuals who have been in a lower than expected grade from delayed school entry or grade repetition may have been subject to the reform before others from the same birth cohort (see [Holmlund, 2007](#)). The resulting attenuation bias can be reduced by dropping all individuals born in the cohort just preceding local implementation of the reform. Fourth, we use the maximum of both parents' (instead of the father's) educational attainment or income. Fifth, we include additional controls for the birth cohort of fathers (first generation) or offspring (second generation estimates). Finally, we include municipality-specific linear time trends to support the common trends assumption that is underlying our difference-in-differences analysis.

Our estimates of the reform effect on the intergenerational educational coefficient remain statistically significant on the $p < 0.001$ level across all specifications. Their sizes vary either very little or as expected. In particular, they increase in absolute size when measurement error in the reform indicator is being addressed (column 4). Estimates differ slightly also when we estimate a parent-offspring (instead of father-offspring) measure of persistence, using maximum education among both mothers and fathers as independent variable (column 5). Estimates of the reform effect on the intergenerational income elasticity have always the same sign, but vary more strongly and are not always statistically significant on the $p < 0.05$ or even $p < 0.1$ level. Two factors reduce precision. First, long-run income is measured with much larger error than educational attainment. Second, the reform had a mechanic and strong effect on the distribution of educational attainment, while incomes were only indirectly affected.

Overall the tests corroborate the existence and the direction of reform effects on the intergenerational persistence in both education and income, but underscore that the former is more precisely estimated. We provide further evidence on the suitability of our identification strategy and the common trends assumption by performing a number of placebo tests. Following Meghir et al. (2011) we falsely assume that the reform took place before or after the actual implementation date. We first sample only those offspring born in 1966 to 1972 whose fathers were subject to the reform and generate a placebo “non-treated” group by pretending that the school reform was implemented one year later, two years, three years, and so on. Similarly, we sample only those fathers who were not treated and pretend that the reform was implemented earlier, thus generating a placebo “treated” group. The resulting estimates are plotted in Figure 2.11.³⁷

Each dot represents the estimate of the reform effect on the intergenerational educational coefficient assuming the reform took place at the specified period before or after the actual implementation date. The largest estimate is obtained when we use the correct timing for the reform assignment (at zero). We find small and insignificant estimates in all other cases, except when we assume that the reform was implemented one year before the actual date. Measurement error in reform status is a potential explanation for this observation, as discussed above and also visible from Figure 2.8 – those in a lower than expected grade may have been subject to the reform even though not captured by our reform indicator (see Holmlund, 2007).

2.4 From Generations to Cohorts

Our model is broadly in line with the previous literature, but motivated by our empirical application we will next relax its coarse generational perspective.³⁸ The existing theoretical literature considers intergenerational transmission between *generations*, but empirical studies estimate mobility trends over *cohorts*. These two dimensions, which do not match due to variation of parental age at birth, have to our knowledge not yet been linked in the literature. We therefore introduce a cohort dimension into our model.

³⁷Corresponding tests provide supportive evidence also for the first-generation estimates (available upon request).

³⁸A more detailed discussion of our theoretical model is given in Nybom and Stuhler (2013), in which we discuss some of its other simplifying assumptions. We demonstrate that our results are not sensitive to the way the influence of parental income is modeled, and that more recursive causal mechanisms (independent effects from grandparents) lead to prolonged dynamic responses of mobility trends to structural shocks.

Our initial motivation was to provide a closer match to the empirical literature, but this extension will also reveal a prospective avenue for identification of past structural changes in mobility levels and trends.

We adopt the following notation to distinguish cohorts and generations. Let the random variable C_t denote the cohort into which a member of generation t of a family is born. Let $A_{t-1,C(t)}$ be a random variable that denotes the age of the parent at birth of the offspring generation t born in cohort C_t . For simplicity assume $A_{t-1,C(t)}$ to be independent of parental income and characteristics, but allow for dependence on C_t , so that its distribution can change over time. Member $t - j$ of a family is then born in cohort

$$C_{t-j} = C_t - A_{t-1,C(t)} - \dots - A_{t-j,C(t-j+1)}. \quad (2.29)$$

Denote realizations of these random variables by lower case letters. For simplicity we consider the scalar case with a single skill. Our reduced two-equations model for intergenerational transmission between offspring born into cohort $C_t = c_t$ and a parent born in cohort $C_{t-1} = c_{t-1}$ is then given by

$$y_{t,c(t)} = \gamma_{c(t)} y_{t-1,c(t-1)} + \rho_{c(t)} e_{t,c(t)} + u_{t,c(t)} \quad (2.30)$$

$$e_{t,c(t)} = \lambda_{c(t)} e_{t-1,c(t-1)} + v_{t,c(t)}, \quad (2.31)$$

where we keep the simplifying assumptions as in our baseline model in equations (2.5) and (2.6). By considering a single set of equations for each generation we abstract from life-cycle effects within a given generation. The transmission parameters in (2.30) and (2.31) can thus be interpreted as representing an average of effective transmission mechanisms over the life-cycle. For example, the price parameter $\rho_{c(t)}$ reflects average returns throughout the working life of an individual born in year c_t .³⁹

Consider for simplicity again the special case in which cross-sectional inequality remains constant, such that $Var(y_{t,c(t)}) = Var(e_{t,c(t)}) = 1 \forall t, c(t)$. Using (2.30) and (2.31), the intergenerational income elasticity of the offspring generation t born in cohort

³⁹A consideration of life-cycle effects (as in [Conlisk, 1969](#), or [Cunha and Heckman, 2007](#)) would be interesting, but the general implications that we discuss here hold as long as some intergenerational transmission mechanisms tend to be effective in early life (e.g., genetic transmission, childhood environment, and education).

c_t then equals

$$\beta_{t,c(t)} = \frac{Cov(y_{t,c(t)}, y_{t-1,C(t-1)})}{Var(y_{t,c(t)})} = \gamma_{c(t)} + \rho_{c(t)} \lambda_{c(t)} Cov(e_{t-1,C(t-1)}, y_{t-1,C(t-1)}), \quad (2.32)$$

where for convenience we do not explicitly condition on $C_t = c_t$. Mobility for a given cohort depends on cohort-specific transmission mechanisms and the covariance of income and endowments in the parent generation. However, this cross-covariance may vary with parental age, since different parental cohorts might have been subject to different policies and institutions. Using eq. (2.29) and the law of iterated expectations we rewrite eq. (2.32) as

$$\begin{aligned} \beta_{t,c(t)} &= \gamma_{c(t)} + \rho_{c(t)} \lambda_{c(t)} E_{A(t-1)} (Cov(e_{t-1,c(t)-A(t-1)}, y_{t-1,c(t)-A(t-1)} | A_{t-1,c(t)})) \\ &= \gamma_{c(t)} + \rho_{c(t)} \lambda_{c(t)} \sum_{a_{t-1}} f_{c(t)}(a_{t-1}) Cov(e_{t-1,c(t)-a(t-1)}, y_{t-1,c(t)-a(t-1)}), \end{aligned} \quad (2.33)$$

where $f_{c(t)}$ is the probability mass function for parental age at birth of cohort c_t . Income mobility thus depends on current transmission mechanisms and a *weighted average* of the cross-covariance of income and endowments in previous cohorts, where the weights are given by the cohort-specific distribution of parental age in the population.⁴⁰

We can iterate backwards to express $\beta_{t,c(t)}$ in terms of parameter values only, and find

$$\beta_{t,c(t)} = \gamma_{c(t)} + \rho_{c(t)} \lambda_{c(t)} \sum_{a_{t-1}} f_{c(t)}(a_{t-1}) \rho_{c(t)-a(t-1)} + \rho_{c(t)} \lambda_{c(t)} \sum_{r=1}^{\infty} z_r, \quad (2.34)$$

where

$$z_r = \sum_{a_{t-1}} \left(f_{c(t)}(a_{t-1}) \cdots \sum_{a_{t-r-1}} \left(f_{c(t-r)}(a_{t-r-1}) \prod_{s=1}^r \left(\gamma_{c(t-s)} \lambda_{c(t-s)} \right) \rho_{c(t-r-1)} \right) \right).$$

Equation (2.34) summarizes how mobility trends across cohorts respond to structural changes. The insights from the generations-only model still hold, but the explicit consideration of cohorts leads to a number of additional implications.⁴¹

⁴⁰The decomposition of the cross-covariance of income and endowments into conditional cross-covariances was simplified here by assuming that first moments of the distribution of those variables are constant over cohorts. In the empirical application we consider cases in which those moments are not constant.

⁴¹In steady state, both equations (2.8) and (2.34) simplify to equation (2.9). The explicit consideration of cohorts has consequences only for transitions between steady states, which may explain why existing

First, while a rapid structural change may have a sudden impact on mobility in the first generation, their effect on mobility trends in subsequent generations will be *gradual* due to variation of parental age at birth. This is exactly the pattern we found in our empirical application (see Figures 2.5 and 2.7).

Second, the importance of past institutions and policies on current mobility rises with parental age at birth. Likewise, the impact of structural changes on mobility trends will die out faster in populations in which individuals become parents at younger ages. Cross-country mobility differentials are thus not only driven by differences in both current and past transmission mechanisms, but also by different *weights* on past mechanisms. This argument might be particularly relevant for comparisons between developed and developing countries.⁴²

Finally, equation (2.34) points to a potential avenue for identification of *past* structural changes in *current* mobility trends, exploiting that the influence of the former on the latter is a function of parental age at birth. As an example, assume that from cohort c^* onwards an expansion of public childcare reduces the heritability of endowments from λ_1 to λ_2 .⁴³ Assume that all parents of generation t were not yet subject to the new regime, such that

$$\lambda_{C(t-1)} = \begin{cases} \lambda_1 & \text{for } C_{t-1} < c^* \\ \lambda_2 & \text{for } C_{t-1} \geq c^* \end{cases}.$$

Other parameters remain unchanged and all grandparents have been subject to the old regime. From equation (2.34), the *conditional intergenerational elasticities* among children with old ($C_{t-1} < c^*$) or young ($C_{t-1} \geq c^*$) parents equal

$$\beta_{t,c(t)} \Big|_{C_{t-1} < c^*} = \gamma + \rho \lambda_2 \gamma \lambda_1 \text{Cov}(e_{t-2,C(t-2)}, y_{t-2,C(t-2)}) + \rho^2 \lambda_1 \quad (2.35)$$

and

$$\beta_{t,c(t)} \Big|_{C_{t-1} \geq c^*} = \gamma + \rho \lambda_2 \gamma \lambda_2 \text{Cov}(e_{t-2,C(t-2)}, y_{t-2,C(t-2)}) + \rho^2 \lambda_1. \quad (2.36)$$

steady-state models have not yet been explicitly linked to cohort-specific measures of mobility.

⁴²Our results imply that mobility in developing countries, in which parents tend to be younger, is less dependent on past institutions. Our example in Section (2.2.3) points to another potential source for *high* mobility in developing countries, in which returns to certain skills or regional wage levels may be comparatively variable over time (e.g., due to internal conflict or rapid economic and societal change).

⁴³For example, [Havnes and Mogstad \(2012\)](#) find that access to subsidized childcare in Norway benefited children from low-income parents the most.

Differencing equations (2.35) and (2.36) then reveals the dynamic, or second-generation impact of the reform on current mobility levels. In practice we may of course encounter various obstacles that are ignored in this simple example. In particular, parental age is likely to correlate with other parental characteristics and thus mobility of their offspring.

A more targeted analysis was feasible in our empirical application: we directly conditioned on parental exposure to a particular school reform, exploiting our knowledge of the geographic variation in its time of effectiveness. We can use the same application to illustrate that even in the absence of such direct evidence, a comparison of conditional mobility measures may still provide a first clue about dynamic effects of past events on current trends. Panel (A) in Figure 2.12 plots conditional coefficients from a regression of offspring on father's years of schooling, for cohorts born from the 1960s until 1972. The pattern is consistent with our previous results: the intergenerational coefficient increases first among families with younger fathers, who were more likely to have been subject to the school reform themselves. Panel (B) shows that the corresponding trend in the intergenerational correlation coefficient is not systematically related to parental age at birth.

2.5 Conclusions

We examined the dynamic relationship between intergenerational mobility in economic outcomes and its underlying structural factors. We showed, theoretically and empirically, that changes in the economic environment affect intergenerational persistence not only in directly affected but also in *subsequent* generations.

Our objective in the empirical application was to identify such dynamic effects for a particular policy reform. Using administrative microdata over three generations, we showed that a Swedish compulsory schooling reform decreased educational and income persistence in directly affected cohorts – by up to a fourth among earlier cohorts, in which the compulsory requirement affected a larger share of the population. But the reform's impact in the subsequent generation was of comparable magnitude, increasing the intergenerational educational coefficient and income elasticity and thus lowering mobility. This second-generation effect is likely to extend to very recent cohorts, as the majority of parents who were themselves subject to the reform had not yet had children when our sample ends. By looking solely at directly affected cohorts, previous research

on similar reforms has thus likely overstated their long-run (or net) mobility effects.

We based our theoretical analysis on a simple simultaneous-equations model, deviating from the existing literature in our focus on its dynamic properties and our consideration of a multidimensional skill vector. We showed that mobility today depends not only on current transmission mechanisms, but also on the joint distribution of income and endowments in past generations – and thus on past mechanisms. Policy or institutional reforms generate therefore long-lasting mobility trends, which are often non-monotonic. Some implications may be surprising, especially our finding that negative mobility trends today can stem from *gains* in equality of opportunity in the past. Other conclusions may have a more intuitive appeal, such that mobility will tend to be higher in times of structural changes.

While the focus was on the general relationship between causal transmission mechanisms and mobility trends, we also noted various practical implications. For example, we showed that the impact of rising wage differentials in US and other countries on mobility may not yet have been fully realized in current data. Changing returns to skills shift intergenerational mobility over at least two generations, while other measures of persistence respond more immediately. This argument may explain why the empirical literature finds increasing sibling correlations in earnings in the US, but less evidence for a corresponding increase in intergenerational persistence. The latter has been surprising as both theoretical (Solon, 2004) and cross-country evidence (e.g., Corak, 2013) suggest a negative relation between cross-sectional inequality and intergenerational mobility.

This implication may be of concern for mobility proponents, as it suggests that a recent decline in mobility might yet to be uncovered by empirical research. But our results also point to a rather innocuous explanation for such observation. We showed that a shift towards a more meritocratic society (a rise in the importance of own skill relative to parental status) tends to generate a non-monotonic response – a mobility gain in the first affected generation, followed by a long-lasting negative trend. We should then perhaps expect mobility to decline in countries that became more meritocratic and mobile in the first half of the 20th century.

Finally, our finding that intergenerational mobility tends to be high in times of change seems consistent with recent evidence from the empirical literature. Long and Ferrie (2013b) find that US occupational mobility was comparatively high in the late 19th century, and suggest that an exceptional degree of geographic mobility may have raised

intergenerational mobility. Our model points to a potential joint cause for both: strong variation in economic conditions across areas and time not only incentivizes internal migration, it also increases intergenerational mobility by altering the local demand conditions that parents and children face during their lifetimes.

Our model is of course highly stylized, and a thorough discussion of related applications requires careful treatment of issues that we only touched upon (such as the timing of intergenerational transmission over an individual's life-cycle, or the difficulties that hinder reliable estimation of trends in income mobility). We briefly addressed promising avenues for future empirical research, noting that different potential causes of mobility shifts could be distinguished by their divergent dynamic implications; that the covariance between income and endowments in the *parent* generation plays a central role in the evolution of income mobility over generations; and that estimation of mobility measures conditional on parental age at birth may provide initial evidence on the effect of past events on current mobility trends.

Tables and Figures

Table 2.1: Sample Statistics by Birth Cohort

	Source data			Intergenerational samples				
	# obs.	reform shares (offspring)	reform shares (fathers)	# obs.	with non-missing (educ.)	(inc.)	reform shares (offspring)	reform shares (fathers)
1943	42,138	0.04	0.00	17,211	15,008	11,059	0.04	0.00
1944	44,715	0.06	0.00	18,425	16,179	14,016	0.06	0.00
1945	44,682	0.06	0.00	18,604	16,441	15,984	0.07	0.00
1946	44,299	0.11	0.00	19,124	17,101	16,800	0.11	0.00
1947	43,288	0.18	0.00	19,078	17,103	16,775	0.18	0.00
1948	42,527	0.31	0.00	19,063	17,192	16,881	0.31	0.00
1949	40,628	0.39	0.00	18,449	16,768	16,424	0.40	0.00
1950	38,854	0.53	0.00	19,421	17,657	17,288	0.54	0.00
1951	36,951	0.56	0.00	18,644	17,016	16,693	0.57	0.00
1952	37,031	0.69	0.00	19,102	17,442	17,085	0.70	0.00
1953	37,537	0.79	0.00	19,452	17,904	17,565	0.80	0.00
1954	35,668	0.86	0.00	18,453	16,955	16,589	0.87	0.00
1955	36,440	0.95	0.00	19,122	17,569	17,179	0.96	0.00
1956	36,666	1.00	0.00	20,942	19,217	18,714	1.00	0.00
...
1965	42,909	1.00	0.01	28,447	26,762	24,657	1.00	0.01
1966	43,050	1.00	0.01	29,043	27,415	25,166	1.00	0.02
1967	42,686	1.00	0.02	28,897	27,366	25,177	1.00	0.03
1968	54,105	1.00	0.04	33,526	32,524	30,124	1.00	0.05
1969	52,317	1.00	0.05	32,157	31,315	28,924	1.00	0.06
1970	53,908	1.00	0.07	32,508	31,788	29,195	1.00	0.08
1971	56,493	1.00	0.09	33,251	32,539	29,783	1.00	0.12
1972	57,035	1.00	0.12	33,081	32,409	29,472	1.00	0.16

Note: Father-child pairs are included in the intergenerational sample if father's age at birth of the child is below

33.

Table 2.2: Reform Effect on Educational and Income Mobility, Cohorts 1943-1955

Panel A: Education		education offspring (# years)		
	(1)	(2)	(3)	(4)
education father (# years)	0.359*** (0.00383)	0.396*** (0.00496)	0.454*** (0.0233)	0.422*** (0.00750)
reform		1.407*** (0.0577)	0.977*** (0.0696)	0.555*** (0.0672)
reform x education father		-0.0969*** (0.00632)	-0.0639*** (0.00685)	-0.0371*** (0.00722)
constant	8.331*** (0.0433)	7.770*** (0.0477)	7.298*** (0.216)	7.306*** (0.0683)
<i>N</i>	220335	220335	220335	220335
Panel B: Income		log income offspring		
	(1)	(2)	(3)	(4)
log inc. father	0.164*** (0.00265)	0.157*** (0.00402)	0.172*** (0.0194)	0.139*** (0.0162)
reform		-0.0111 (0.0759)	0.102 (0.0936)	0.253** (0.121)
reform x log inc. father		0.00510 (0.00618)	-0.00588 (0.00760)	-0.0196** (0.00995)
constant	9.893*** (0.0324)	9.947*** (0.0487)	9.762*** (0.236)	9.915*** (0.195)
<i>N</i>	199340	199340	199340	199340
municipality controls				x
offspring cohort controls			x	x

Note: Clustered (municipality level) standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Coefficient estimates from equation (2.24) (column 4) and simplified variants (columns 1-3), based on offspring cohorts 1943-1955 in intergenerational sample.

Table 2.3: Reform Effect on Educational and Income Mobility, Cohorts 1966-1972

Panel A: Education		education offspring (# years)			
	(1)	(2)	(3)	(4)	
education father (# years)	0.240*** (0.00214)	0.238*** (0.00298)	0.195*** (0.00929)	0.294*** (0.00411)	
reform (father)		-0.904*** (0.0894)	-0.923*** (0.0847)	-0.768*** (0.139)	
reform x education father		0.0534*** (0.00893)	0.0727*** (0.00762)	0.0655*** (0.0128)	
constant	9.741*** (0.0233)	9.813*** (0.0356)	10.07*** (0.0938)	9.763*** (0.0471)	
<i>N</i>	111173	111173	111173	111173	
Panel B: Income		log income offspring			
	(1)	(2)	(3)	(4)	
log inc. father	0.207*** (0.00530)	0.211*** (0.00601)	0.186*** (0.0172)	0.244*** (0.0093)	
reform (father)		0.331** (0.156)	-0.0949 (0.164)	-0.498* (0.265)	
reform x log inc. father		-0.0286** (0.0128)	0.00814 (0.0134)	0.0410* (0.0216)	
constant	9.666*** (0.0650)	9.618*** (0.0734)	9.874*** (0.210)	9.446*** (0.118)	
<i>N</i>	110317	110317	110317	110317	
municipality controls				x	
father cohort controls			x	x	

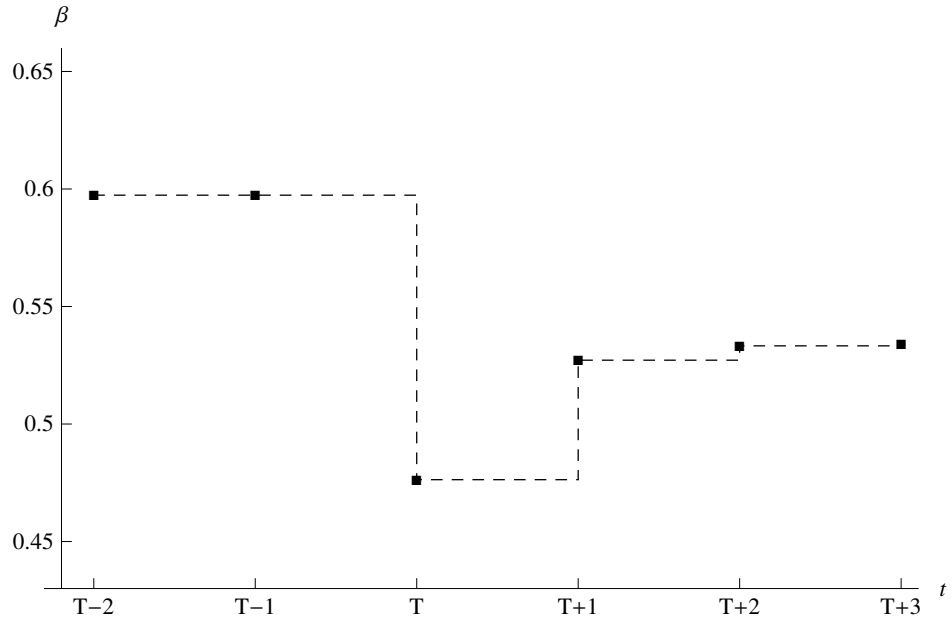
Note: Clustered (municipality level) standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Coefficient estimates from equation (2.28) (column 4) and simplified variants (columns 1-3), based on offspring cohorts 1966-1972 in intergenerational sample.

Table 2.4: Robustness Tests

	baseline	with siblings	fathers below 30	pre-reform dropped	parental max.	cohort controls	municip. time trends
Education:							
1st gen.	-0.0371*** (0.0072)	-0.0393*** (0.0054)	-0.0408*** (0.0089)	-0.0434*** (0.0083)	-0.0357*** (0.0064)	-0.0387*** (0.0073)	-0.0364*** (0.0074)
2nd gen.	0.0655*** (0.0128)	0.0651*** (0.0122)	0.0655*** (0.0128)	0.0710*** (0.0139)	0.0307*** (0.0093)	0.0655*** (0.0126)	0.0622*** (0.0131)
Income:							
1st gen.	-0.0196* (0.0100)	-0.0078 (0.0068)	-0.0181 (0.0115)	-0.0195* (0.0118)	-0.0210** (0.0088)	-0.0233** (0.0095)	-0.0239** (0.0097)
2nd gen.	0.0410* (0.0216)	0.0148 (0.0165)	0.0410* (0.0216)	0.0492** (0.0238)	0.0344** (0.0155)	0.0418** (0.0212)	0.0363* (0.0219)

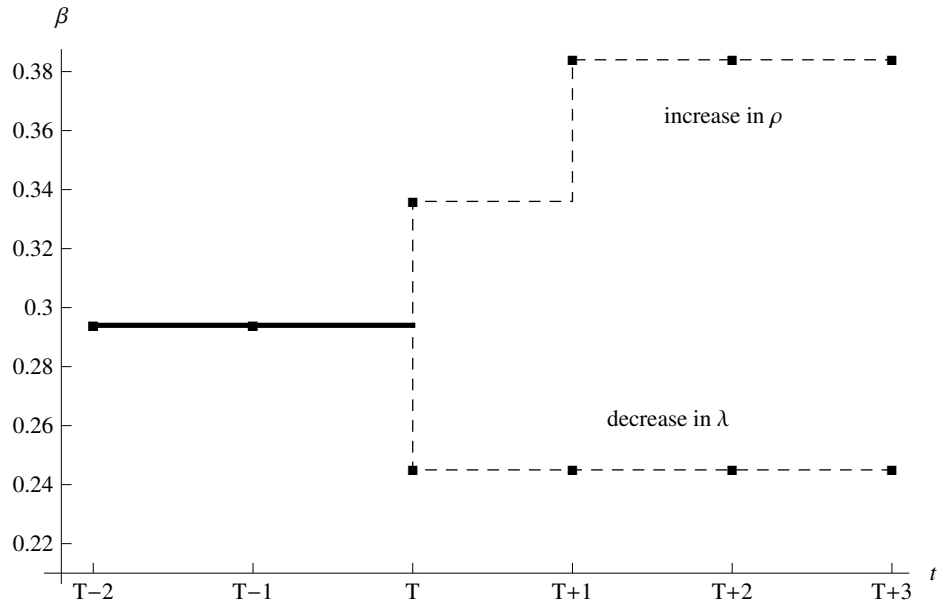
Note: Sensitivity analyses reporting the coefficient on the interaction between reform dummy and parental education and income and clustered standard errors (in parentheses), * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Column 1 contains the baseline specification. For the next columns we include the sibling subsample, restrict the sample to fathers with age at birth below 30, drop offspring born in the cohort preceeding the reform implementation, use the maximum of mother's and father's education or income, include father (rows 1 and 3) or offspring cohort dummies (rows 5 and 7), or include municipality-specific linear trends.

Figure 2.2: A Declining Impact of Parental Income and Increasing Returns to Skills



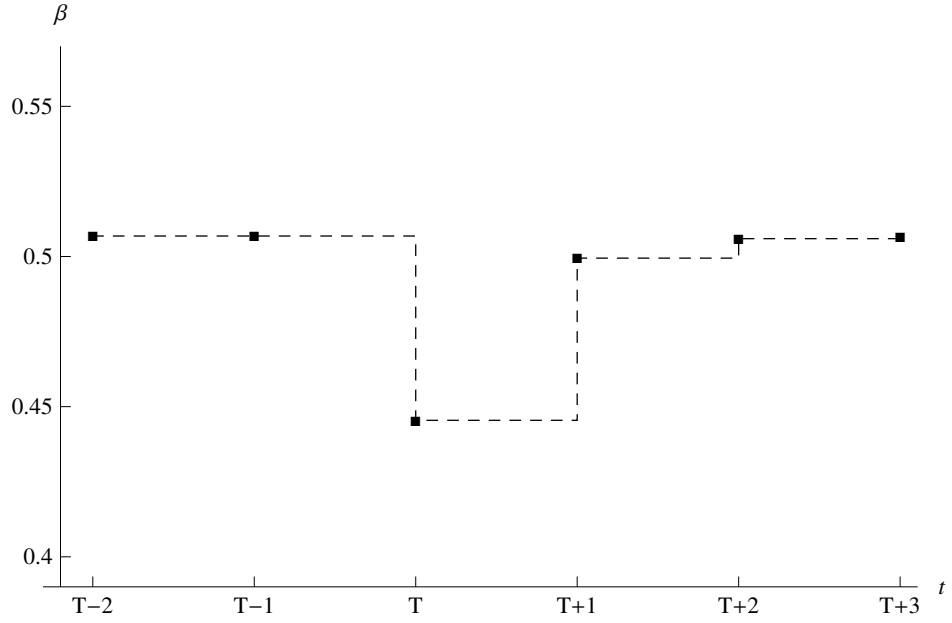
Note: Mobility trend over generations in numerical example. In generation T the impact of parental income γ declines from $\gamma_1 = 0.4$ to $\gamma_2 = 0.2$ while the returns to endowments and human capital ρ increase from $\rho_1 = 0.5$ to $\rho_2 = 0.7$ (assuming $\lambda = 0.6$).

Figure 2.1: A Change in the Heritability of, or Returns to, Endowments



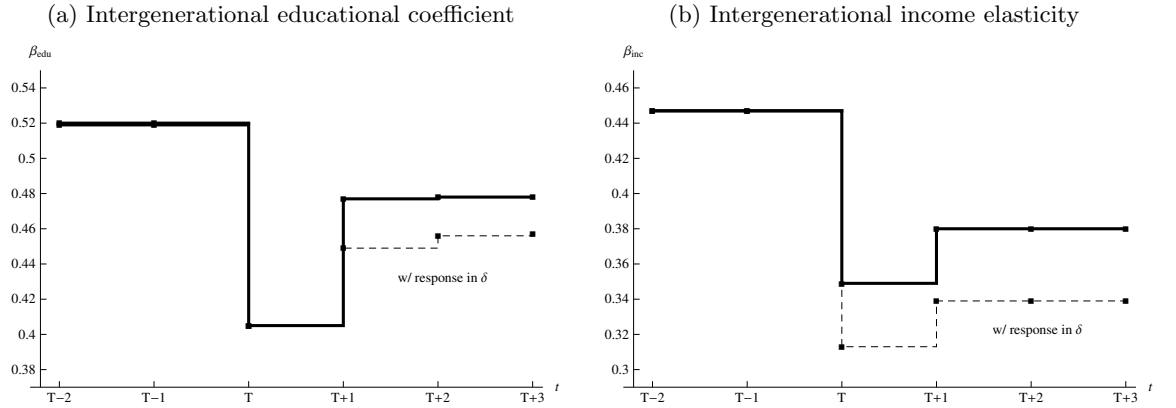
Note: Mobility trend over generations in two numerical examples. Example 1a: in generation T the heritability of endowments λ decreases from $\lambda_1 = 0.6$ to $\lambda_2 = 0.5$ (assuming $\rho = 0.7$ and $\gamma = 0$). Example 1b: the returns to endowments and human capital ρ increase from $\rho_1 = 0.7$ to $\rho_2 = 0.8$ (assuming $\lambda = 0.6$).

Figure 2.3: A Swap in Prices



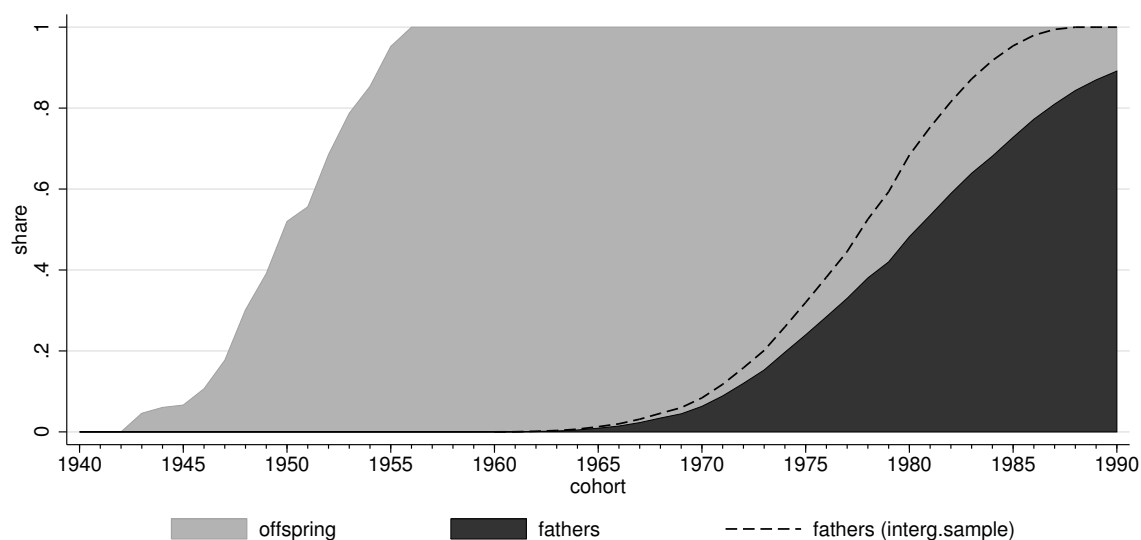
Note: Mobility trend over generations in numerical example. In generation T the returns to skill k increase from $\rho_{k,1} = 0.3$ to $\rho_{k,2} = 0.6$ and the returns to skill l decrease from $\rho_{l,1} = 0.6$ to $\rho_{l,2} = 0.3$ (assuming $\gamma = 0.2$ and $\lambda = 0.6$).

Figure 2.4: Raising the Compulsory Schooling Level



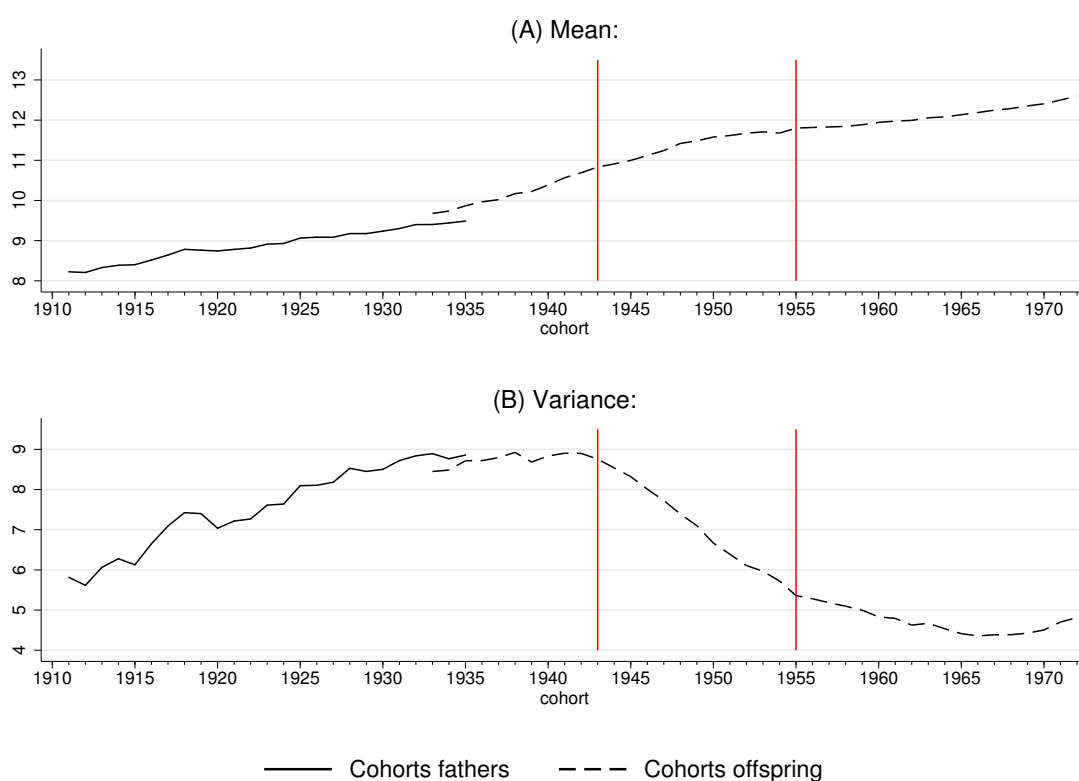
Note: Income and educational mobility trends in numerical example, with $x = 9$, $\alpha_y = 9$, $\gamma_y = 0$, $\delta = 0.2$ (dashed line: $\delta = 0.18$), $\alpha_h = 10$, $\gamma_h = 1$, $\theta = 2$, $\lambda = 0.6$, and (u_y, u_h, v) normally distributed with variances $(0.1, 2.75, 0.64)$.

Figure 2.5: Share of Offspring and Fathers Subject to Reform



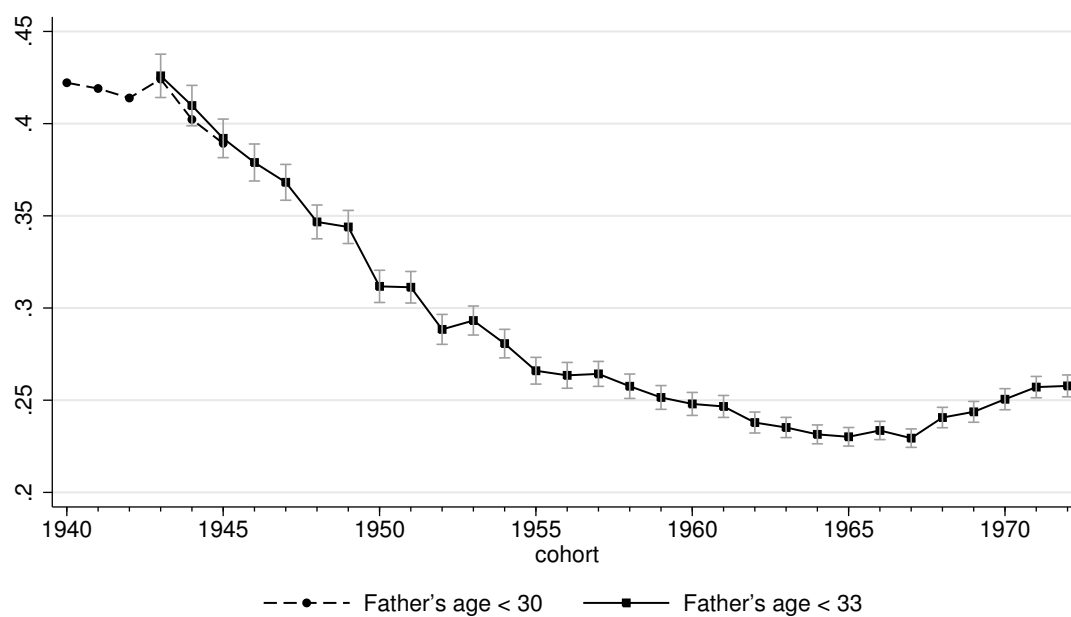
Note: Share of offspring and fathers subject to school reform over offspring cohorts, in source data (grey and black areas) and intergenerational sample (dashed line).

Figure 2.6: Mean and Variance of Years of Schooling over Cohorts



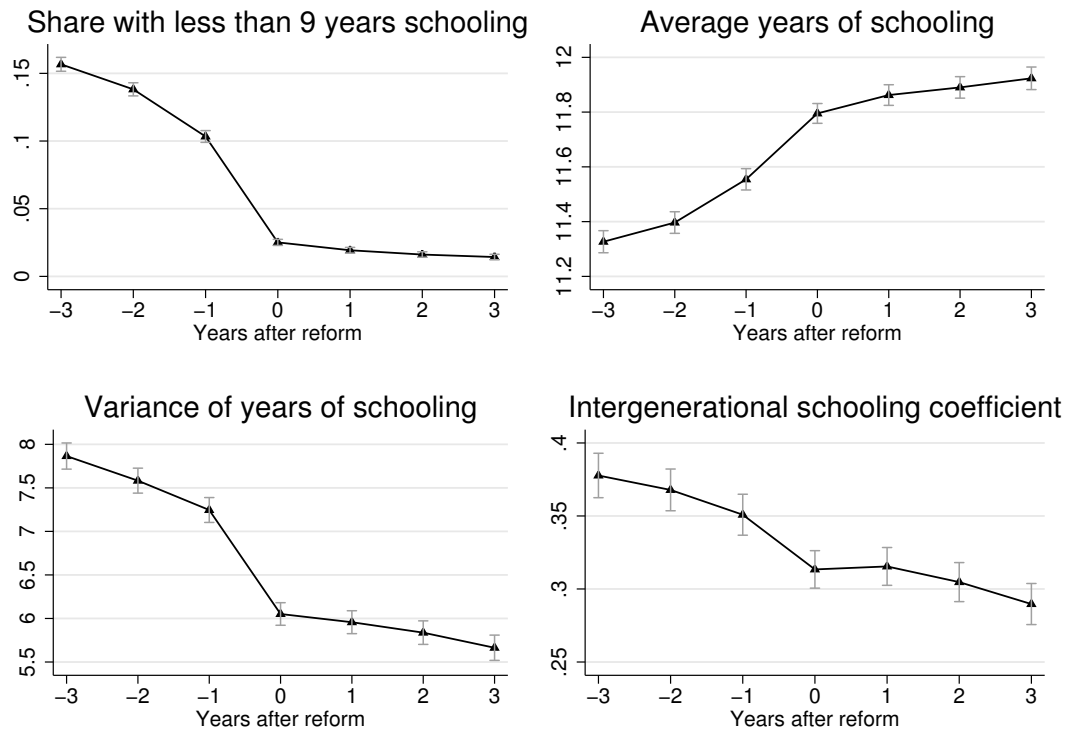
Note: Moments of years of schooling over cohorts of offspring (dashed line) and their fathers (solid line) in intergenerational sample.

Figure 2.7: Trends in the Intergenerational Educational Coefficient over Cohorts



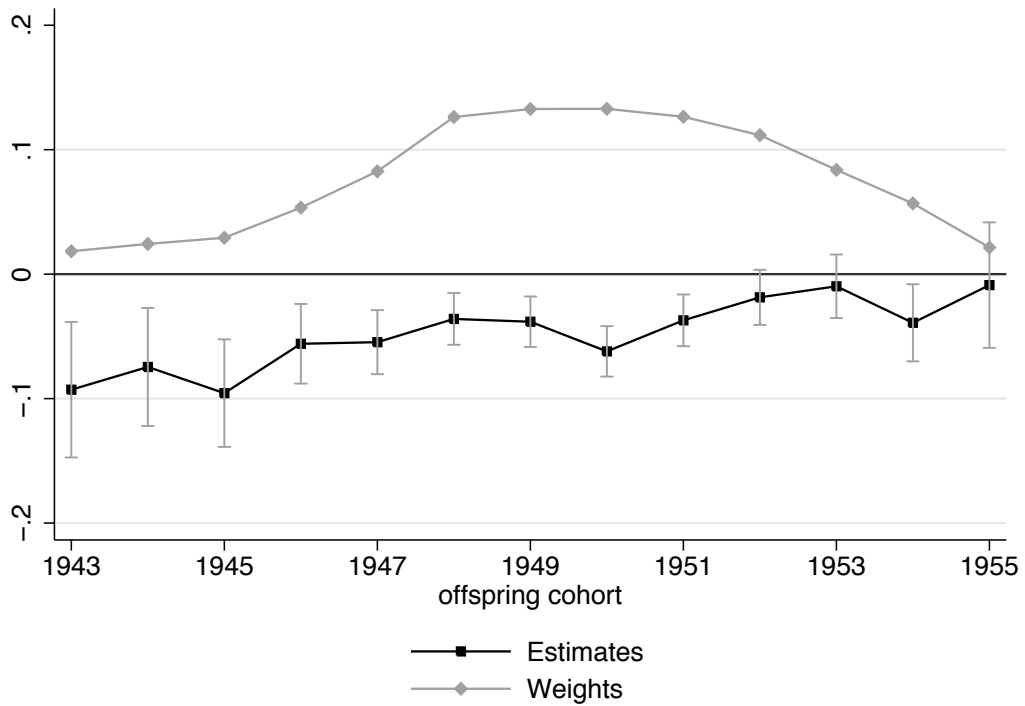
Note: Each dot represents the coefficient from a regression of years of schooling of offspring in the respective birth cohort on years of schooling of their fathers. Based on intergenerational sample (fathers aged below 33, solid line) and subsample (fathers aged below 30, dashed line). Grey bars: 95% confidence intervals.

Figure 2.8: Educational Attainment and Mobility, Pre- vs. Post-Reform



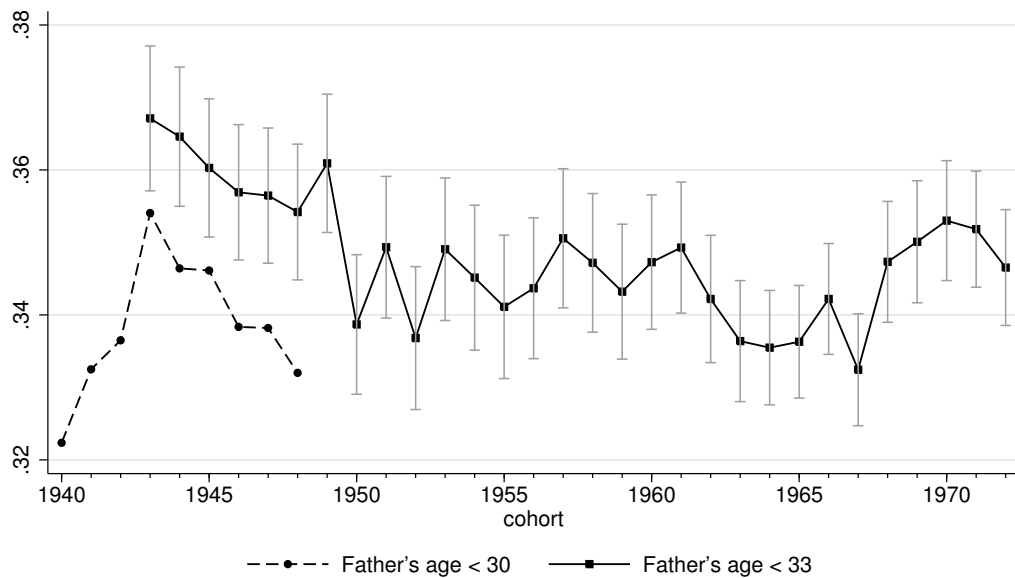
Note: We recenter the data such that the reform occurs at time zero for each municipality. Panels (a)-(c) summarize the distribution of offspring educational attainment. Each dot in panel (d) represents the coefficient from a regression of years of schooling of offspring on years of schooling of their fathers. Based on intergenerational sample (fathers aged below 33). Grey bars: 95% confidence intervals.

Figure 2.9: Heterogeneity in the Reform Effect over Cohorts



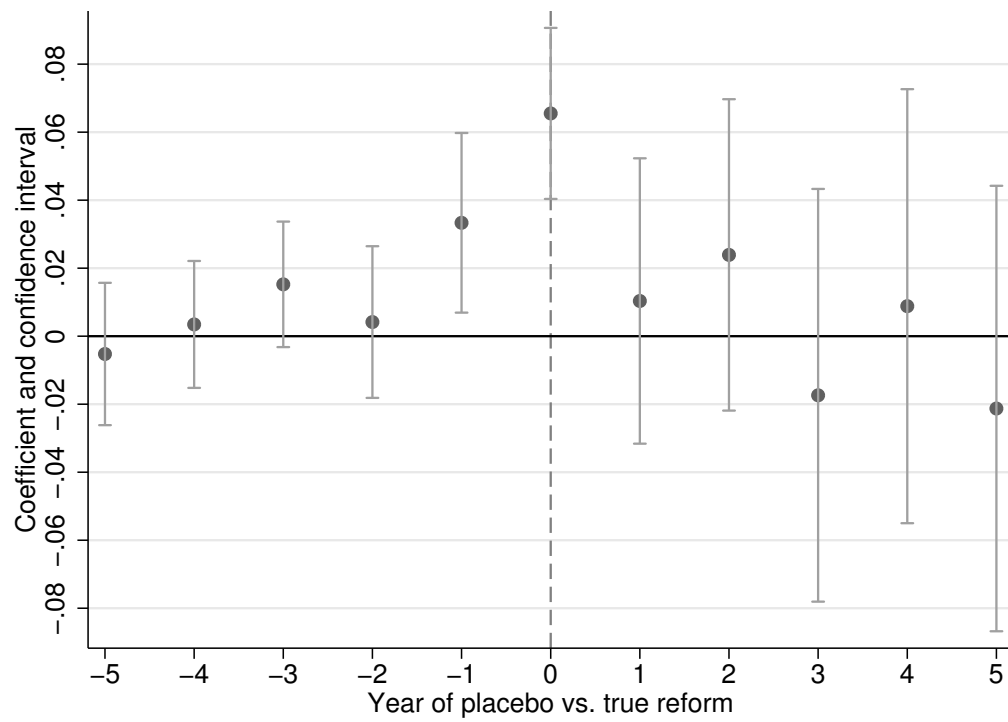
Note: Estimates of the reform effect on the intergenerational educational coefficient over cohorts (black line), and their respective weight in the pooled coefficient (grey line). Based on intergenerational sample (fathers aged below 33), including sibling subsample. Grey bars: 95% confidence intervals.

Figure 2.10: Trends in the Intergenerational Educational Correlation over Cohorts



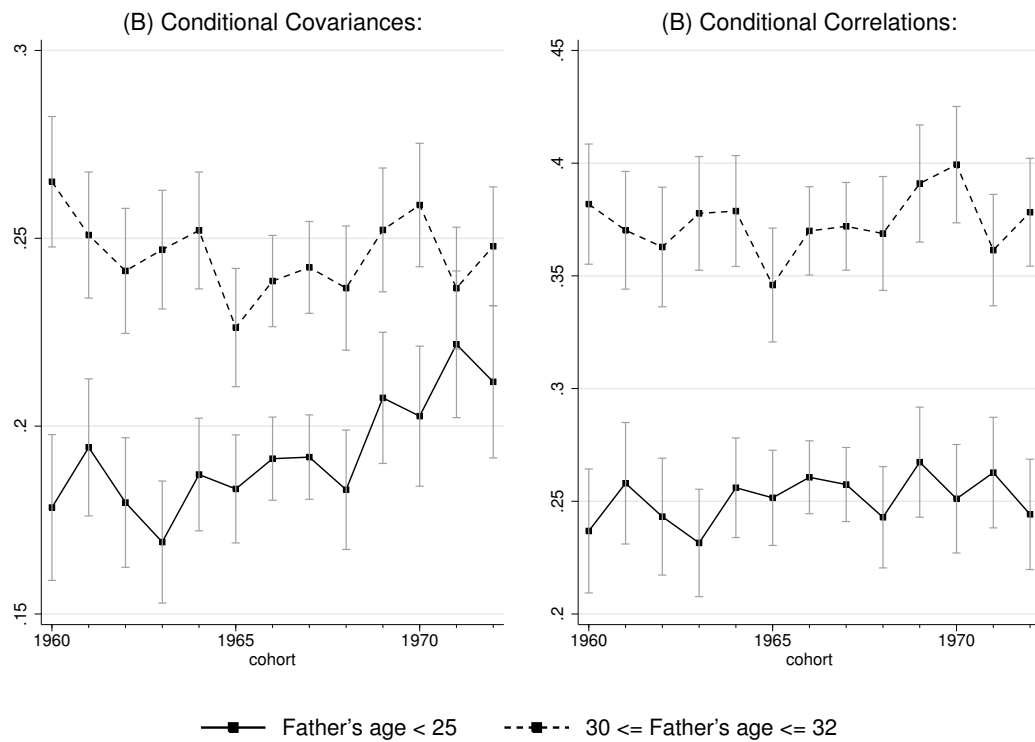
Note: Each dot represents the correlation coefficient between years of schooling of offspring in the respective birth cohort and years of schooling of their fathers. Based on intergenerational sample (fathers aged below 33, solid line) and subsample (fathers aged below 30, dashed line). Grey bars: 95% confidence intervals.

Figure 2.11: Placebo Test: Second Generation



Note: Each dot represents an estimate of the reform effect on the intergenerational educational coefficient in cohorts 1966-72 under the assumption that the reform took place at the specified period before or after the actual implementation date. Based on intergenerational sample (fathers aged below 33). Grey bars: 95% confidence intervals.

Figure 2.12: Trends in Conditional Intergenerational Regression and Correlation Coefficients



Note: Each dot in panel (A) represents the coefficient from a regression of offspring years of schooling in the respective birth cohort on father years of schooling, for fathers aged 18-24 (solid line) or fathers aged 30-32 (dashed line) at offspring birth. Panel (B) presents the corresponding correlation coefficients. Grey bars: 95% confidence intervals.

Appendix

A.1 An Economic Model of Intergenerational Transmission

We model the optimizing behavior of parents to derive the “mechanical” transmission equations presented in Section 2.2. For this purpose we extend the model in Solon (2004), considering parental investments in multiple distinctive types of human capital and statistical discrimination on the labor market.

Assume that parents allocate their lifetime *after tax earnings* $(1 - \tau)Y_{t-1}$ between own *consumption* C_{t-1} and *investments* $I_{1,t-1}, \dots, I_{J,t-1}$ in J distinctive types of human capital of their children. Parents do not bequeath financial assets and face the budget constraint

$$(1 - \tau)Y_{t-1} = C_{t-1} + \sum_{j=1}^J I_{j,t-1}. \quad (2.37)$$

Accumulation of *human capital* h of type j in offspring generation t depends on parental investment, a $K \times 1$ vector of inherited *endowments* \mathbf{e}_t , and chance $u_{j,t}$,

$$h_{j,t} = \gamma_j \log I_{j,t-1} + \boldsymbol{\theta}'_j \mathbf{e}_t + u_{j,t} \quad \forall j \in 1, \dots, J, \quad (2.38)$$

where γ_j and elements of the vector $\boldsymbol{\theta}_j$ measure the marginal product of parental investment and each endowment. Endowments represent early child attributes that may be influenced by nature (genetic inheritance) or nurture (e.g. parental upbringing). We assume that they are positively correlated between parents and their children, as implied by the autoregressive process

$$e_{k,t} = \lambda_k e_{k,t-1} + v_{k,t} \quad \forall k \in 1, \dots, K, \quad (2.39)$$

where $v_{k,t}$ is a white-noise error term and the heritability coefficient λ_k lies between 0 and 1. We may allow endowments to be correlated within individuals, leading to the more general transmission equation (2.4). Finally, assume that income of offspring equals

$$\log Y_t = \begin{cases} \boldsymbol{\delta}' \mathbf{h}_t + u_{y,t} & \text{with probability } p \\ \boldsymbol{\delta}' E[\mathbf{h}_t | Y_{t-1}] + u_{y,t} & \text{with probability } 1 - p \end{cases}. \quad (2.40)$$

With probability p employers observe human capital of workers and pay them their

marginal product $\delta' \mathbf{h}_t$ plus a white-noise error term $u_{y,t}$, which reflects market luck. With probability $1 - p$ employers cannot uncover true productivity, and remunerate workers instead for their *expected* productivity given observed parental background. In particular, employers observe that on average parents invest income share s_j in offspring human capital of type j , such that $E[I_{j,t-1}|Y_{t-1}] = s_j Y_{t-1}$, and that the offspring of high-income parents tend to have more favorable endowments, such that $E[e_{k,t}|Y_{t-1}] = \gamma_k Y_{t-1}$ (with $\gamma_k \geq 0$) for all $k \in 1, \dots, K$.

Parents choose investment in the child's human capital as to maximize the utility function

$$U_{t-1} = (1 - \alpha) \log C_{t-1} + \alpha E[\log Y_t | Y_{t-1}, \mathbf{I}_{t-1}, \mathbf{e}_t], \quad (2.41)$$

where the altruism parameter $\alpha \in [0, 1]$ measures the parent's taste for own consumption relative to the child's expected income. Given equations (2.37) to (2.41), the Lagrangian for parent's investment decision is

$$\begin{aligned} \mathcal{L}(C_{t-1}, \mathbf{I}_{t-1}, \mu) &= (1 - \alpha) \log C_{t-1} + \alpha \delta' (p E[\mathbf{h}_t | Y_{t-1}, \mathbf{I}_{t-1}, \mathbf{e}_t] + (1 - p) E[\mathbf{h}_t | Y_{t-1}]) \\ &\quad + \mu ((1 - \tau) Y_{t-1} - C_{t-1} - \mathbf{1}' \mathbf{I}_{t-1}) \end{aligned}$$

The first-order conditions require that

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial C_{t-1}} &= \frac{1 - \alpha}{C_{t-1}} - \mu = 0, \\ \frac{\partial \mathcal{L}}{\partial I_{j,t-1}} &= \frac{\alpha (1 - p) \delta_j \gamma_j}{I_{j,t-1}} - \mu = 0 \quad \forall j \in 1, \dots, J, \\ \frac{\partial \mathcal{L}}{\partial \mu} &= (1 - \tau) Y_{t-1} - C_{t-1} - \mathbf{1}' \mathbf{I}_{t-1} = 0. \end{aligned}$$

Optimal investments,

$$I_{j,t-1} = \frac{\alpha p \delta_j \gamma_j}{(1 - \alpha) + \sum_{l=1}^J \alpha p \delta_l \gamma_l} (1 - \tau) Y_{t-1} \quad \forall j \in 1, \dots, J, \quad (2.42)$$

increase in parental altruism and income, and in the probability that offspring human capital is observed and acted on by employers. Parents invest more into those skills in which the marginal product of investment or the return on the labor market are large. Plugging optimal investment into equation (2.38) yields (ignoring constants, which are irrelevant for our analysis) equation (2.3), which if plugged in turn into eq. (2.40) motivates equation (2.2).

A.2 Reduced Form and Stability

The reduced form of equations (2.5) and (2.6) is

$$\begin{pmatrix} y_t \\ e_t \end{pmatrix} = \begin{pmatrix} \gamma_{y,t} + \delta'_t \gamma_{h,t} & \delta'_t \Theta_t \Lambda_t \\ \mathbf{0} & \Lambda_t \end{pmatrix} \begin{pmatrix} y_{t-1} \\ e_{t-1} \end{pmatrix} + \begin{pmatrix} u_{y,t} + \delta'_t \mathbf{u}_{h,t} + \delta'_t \Theta_t \mathbf{v}_t \\ \mathbf{v}_t \end{pmatrix}, \quad (2.43)$$

which we may shorten to

$$\mathbf{x}_t = \mathbf{A}_t \mathbf{x}_{t-1} + \mathbf{w}_t. \quad (2.44)$$

Let subscripts 1, 2 index parameter values before and after a structural shock occurs in generation T .⁴⁴ The stability condition $\lim_{s \rightarrow \infty} \mathbf{A}_2^s = \mathbf{0}$ is then satisfied by assuming that $\gamma_{y,2} + \delta'_2 \gamma_{h,2}$ and all eigenvalues of Λ_2 are non-negative and below one.⁴⁵ These conditions also ensure that the transitions of the first and second moments of the distribution of \mathbf{x}_t towards their steady state values are monotonic (see [Jenkins, 1982](#)), a property that however does not extend to the transition path of the intergenerational elasticity, as we discuss in Sections 2.2. Normalization of the variances of y_t and elements of \mathbf{h}_t and \mathbf{e}_t in the initial steady state leads to additional parameter restrictions. Take the covariance of (2.44) and denote the covariance matrices of \mathbf{x}_t and \mathbf{w}_t by \mathbf{S}_t and \mathbf{W}_t , such that

$$\mathbf{S}_t = \mathbf{A}_t \mathbf{S}_{t-1} \mathbf{A}_t' + \mathbf{W}_t.$$

Denote by γ , $\boldsymbol{\rho}$, and Λ the steady-state parameter values before a structural change occurs in generation $t = T$. Note that in steady state $\mathbf{S}_t = \mathbf{S}_{t-1} = \mathbf{S}$, normalize all diagonal elements of \mathbf{S} to one, and solve for the variances of $u_{y,t}$ and elements of $\mathbf{u}_{h,t}$ and \mathbf{v}_t . For example, if Λ_t is diagonal then $\text{Var}(e_{j,t}) = 1 \forall j$ iff $\text{Var}(v_{j,t}) = 1 - \lambda_j^2 \forall j$; the variances are non-negative iff $\lambda_{jj} \leq 1 \forall j$, as is also required for stability of the system.

A.3 Choice of Parameter Values

Our main findings do not rely on specific parameter choices, but our numerical examples will benefit from parametrizations that are consistent with the empirical literature. One

⁴⁴[Conlisk \(1974b\)](#) derives stability conditions in a random coefficients model with repeated shocks.

⁴⁵For example, if Λ_2 is diagonal and elements of the endowment vector \mathbf{e}_t are uncorrelated then the diagonal elements of Λ_2 are required to be strictly between zero and one.

difficulty is that some variables in our model represent broad concepts (e.g., human capital \mathbf{h}_t may include any productive characteristic of an individual), which are only imperfectly captured by data. In addition, the parameters of the model reflect total effects from those variables. While estimates of (intergenerational) correlations and other moments are widely reported, there exists less knowledge about the relative importance of the various underlying causal mechanisms. Although only indicative, we can at least choose parameter values that are consistent with the available evidence.

Lefgren et al. (2012) examine the relative importance of different mechanisms in a transmission framework that is similar to ours. Using imperfect instruments that are differentially correlated with parental human capital and income they estimate that in Sweden the effect from parental income (captured by the parameter γ) explains about a third of the intergenerational elasticity, while parental human capital explains the remaining two thirds. In our model we further distinguish between a direct and indirect (through human capital accumulation) effect from parental income, as captured by the parameters γ_y and γ_h , but the total effect is sufficient for the parameterization of our examples.

The literature provides more guidance on the transmission of physical traits such as height or cognitive and non-cognitive abilities, for which we use the term *endowments*. Common to these are that genetic inheritance is expected to play a relatively important role. From the classic work of Galton to more recent studies the evidence implies intergenerational correlations in the order of magnitude of about 0.3-0.4 when considering one and much higher correlations when considering both parents.⁴⁶ Those estimates may reflect to various degrees not only genetic inheritance but also correlated environmental factors; we capture both in the *heritability* parameter λ (estimates of genetic transmission are then a lower bound), for which values in the range 0.5-0.8 seem reasonable. Note that we use the term “heritability” in a broad sense, while the term refers only to genetic inheritance in the biological literature.

Finally, a reasonable lower-bound estimate of the *returns* ρ to endowments and *human capital* can be approximated by evidence on the explanatory power of earnings equations. Studies that observe richer sets of covariates, including measures of cognitive and non-

⁴⁶For estimates of correlations in measures of cognitive ability, see Bowles and Gintis (2002) and the studies they cite; for measures of both cognitive ability and non-cognitive ability, see Grönqvist et al. (2010).

cognitive ability, typically yield estimates of R^2 in the neighborhood of 0.40.⁴⁷ On the one hand, such estimates are likely to underestimate the explanatory power of (broadly defined) human capital as of imperfect measurement and omitted variables. On the other hand, we want to only capture returns to the component of human capital that is not due to parental income and investment; we capture the latter channel instead in the parameter γ_h (and its contribution to offspring income in γ). In any case, values of ρ in the range of 0.6-0.8 should be at least roughly consistent with the empirical evidence.⁴⁸

These parameter ranges are consistent with recent estimates of the intergenerational income elasticity β in the US, which are typically in the range of 0.45-0.55 (see [Black and Devereux, 2011](#)). Given reliable elasticity estimates we can also cross-validate and potentially narrow down the implied range for the structural parameters of the model. We write each parameter as a function of the others in steady state,

$$\begin{aligned} \beta &= \gamma + \frac{\rho^2 \lambda}{1 - \gamma \lambda} & \gamma &= \frac{\beta \lambda + 1 \pm \sqrt{\beta^2 \lambda^2 - 2\beta \lambda + 4\lambda^2 \rho^2 + 1}}{2\lambda} \\ \rho &= \sqrt{\frac{(\beta - \gamma)(1 - \gamma \lambda)}{\lambda}} & \lambda &= \frac{\beta - \gamma}{\beta \gamma + \rho^2 - \gamma^2}, \end{aligned} \tag{2.45}$$

and plug in the discussed values on the right-hand sides to impute parameter ranges that are consistent with our reading of the empirical literature. Specifically we rule out too high values of λ and ρ as they cause γ to approach zero, to arrive at

$$0.45 \leq \beta \leq 0.55, \quad 0.15 \leq \gamma \leq 0.25, \quad 0.60 \leq \rho \leq 0.70, \quad 0.50 \leq \lambda \leq 0.65.$$

These implied ranges should not be taken literally, but are sufficient to provide a reasonable illustration of the potential quantitative implications of our findings.

A.4 Correlated endowments

We revisit example 3 under the assumption that $\mathbf{\Lambda}_t$ is not diagonal, such that elements of the endowment vector \mathbf{e}_t are potentially correlated. Suppose that at generation T the returns to human capital change from $\boldsymbol{\rho}_1$ to $\boldsymbol{\rho}_2$ but that the steady-state variance of

⁴⁷See for example [Lindqvist and Vestman \(2011\)](#) for Sweden. Fixed-effects models yield higher estimates, although some of the difference may be capturing persistent luck rather than unobserved characteristics.

⁴⁸In the initial steady state we standardize $Var(y) = Var(e) = 1$, such that $R^2 = 0.4$ translates into $\rho \approx 0.63$.

income remains unchanged.

By substituting equation (2.5) for y_{T-1} and income in previous generations we can express the pre-shock elasticity as

$$\beta_{T-1} = Cov(y_{T-1}, y_{T-2}) = \gamma + \boldsymbol{\rho}'_1 Cov(\mathbf{e}_{T-1}, y_{T-2}) = \gamma + \boldsymbol{\rho}'_1 \boldsymbol{\Gamma} \boldsymbol{\rho}_1 \quad (2.46)$$

where

$$\boldsymbol{\Gamma} = \sum_{l=1}^{\infty} \gamma^{l-1} Cov(\mathbf{e}_{T-1}, \mathbf{e}_{T-1-l}) \quad (2.47)$$

is the cross-covariance between the endowment vectors of offspring and parents (if $\gamma = 0$), or a weighted average of the endowment vectors of parents and earlier ancestors ($0 < \gamma < 1$). These cross-covariances measure to what degree each offspring endowment is correlated with the *same* endowment in previous generations (the diagonal elements) and each of the *other* $K - 1$ endowments (the off-diagonal elements). Note that $\boldsymbol{\Gamma}$ does not depend on t if these cross-covariances are in steady state.

We can similarly derive the elasticity in the first affected generation and in the new steady state as

$$\beta_T = \gamma + \boldsymbol{\rho}'_2 \boldsymbol{\Gamma} \boldsymbol{\rho}_1 \quad (2.48)$$

$$\beta_{t \rightarrow \infty} = \gamma + \boldsymbol{\rho}'_2 \boldsymbol{\Gamma} \boldsymbol{\rho}_2. \quad (2.49)$$

The conditions under which a change in skill prices leads to a non-monotonic response in mobility can be easily summarized if the cross-covariances $Cov(\mathbf{e}_{T-1}, \mathbf{e}_{T-j}) \forall j > 1$ are symmetric. Symmetry requires the correlation between offspring endowment k and parent endowment l to be as strong as the correlation between offspring endowment l and parent endowment k , $\forall k, l$. We can then note that

$$\begin{aligned} 2\beta_T &= 2(\gamma + \boldsymbol{\rho}'_2 \boldsymbol{\Gamma} \boldsymbol{\rho}_1) \\ &= \gamma + \boldsymbol{\rho}'_1 \boldsymbol{\Gamma} \boldsymbol{\rho}_1 + (\boldsymbol{\rho}'_2 - \boldsymbol{\rho}'_1) \boldsymbol{\Gamma} \boldsymbol{\rho}_1 + \gamma + \boldsymbol{\rho}'_2 \boldsymbol{\Gamma} \boldsymbol{\rho}_2 + \boldsymbol{\rho}'_2 \boldsymbol{\Gamma} (\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2) \\ &= \beta_{T-1} + \beta_{t \rightarrow \infty} + (\boldsymbol{\rho}'_2 - \boldsymbol{\rho}'_1) \boldsymbol{\Gamma} \boldsymbol{\rho}_1 - \boldsymbol{\rho}'_2 \boldsymbol{\Gamma} (\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1) \\ &= \beta_{T-1} + \beta_{t \rightarrow \infty} - (\boldsymbol{\rho}'_2 - \boldsymbol{\rho}'_1)' \boldsymbol{\Gamma} (\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1), \end{aligned} \quad (2.50)$$

where we expanded and subtracted $\boldsymbol{\rho}'_1$ and $\boldsymbol{\rho}_2$, substituted equations (2.46) and (2.49), and finally took the transpose and used the symmetry of $\boldsymbol{\Gamma}$ to collect all remaining terms

in a quadratic form.

Let S denote the subset of prices that do not change in generation T , and denote by $\mathbf{\Gamma}_S$ and $\mathbf{\Lambda}_S$ the minors of $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$ that are formed by deleting each row and column that correspond to an element in S . The quadratic form $(\boldsymbol{\rho}'_2 - \boldsymbol{\rho}'_1)' \mathbf{\Gamma} (\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1)$ is greater than zero for $\boldsymbol{\rho}_2 \neq \boldsymbol{\rho}_1$ if $\mathbf{\Gamma}_S$ is positive definite. A sufficient condition for $\mathbf{\Gamma}_S$ to be positive definite is diagonality of the heritability matrix $\mathbf{\Lambda}_S$, with positive diagonal elements. More generally, the matrix $\mathbf{\Gamma}_S$ is positive definite if the respective minors of the cross-covariances $Cov(\mathbf{e}_{T-1}, \mathbf{e}_{T-j}) \forall j > 1$ are strictly diagonally dominant. Strict diagonal dominance requires that the correlation between offspring endowment k and parent endowment k is stronger than the sum of its correlation to all other relevant parent endowments $l \neq k, l \in S$ (i.e., offspring are similar instead of dissimilar to their parents).

Price changes then increase intergenerational mobility temporarily (β_T is below *both* the previous steady state β_{T-1} and the new steady state $\beta_{t \rightarrow \infty}$) as long as the steady-state elasticity shifts not too strongly, specifically iff

$$|\beta_{t \rightarrow \infty} - \beta_{T-1}| < (\boldsymbol{\rho}'_2 - \boldsymbol{\rho}'_1) \mathbf{\Lambda} (\mathbf{I} - \gamma \mathbf{\Lambda})^{-1} (\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1). \quad (2.51)$$

Chapter 3

Life-Cycle Bias in Intergenerational Mobility Estimation

Transmission of economic status within families is often measured by the intergenerational elasticity between parents' and children's lifetime income. A large and growing literature has estimated this parameter in order to analyze the extent of intergenerational mobility across countries, groups and time.¹ Unfortunately, the estimates in the early literature suffered greatly from measurement error in lifetime income, and successive methodological improvements led to large-scale corrections.²

While the early estimates were severely attenuated from approximation of lifetime values by noisy single-year income data for parents, [Jenkins \(1987\)](#) identifies systematic deviations of current from lifetime values over the life cycle as an additional source of inconsistency. Evidence by [Haider and Solon \(2006\)](#) and [Grawe \(2006\)](#) suggests that the latter is empirically important. Various refined methods to address such *life-cycle bias* have recently been presented. In particular, Haider and Solon proposed a tractable generalization of the classical errors-in-variables model that, while applicable also in other contexts, has strongly influenced how researchers make use of short-run income data in the intergenerational mobility literature.

¹See [Solon \(1999\)](#) for a comprehensive evaluation of the early empirical literature. Recent surveys include [Björklund and Jäntti \(2009\)](#) and [Black and Devereux \(2011\)](#).

²For example, the intergenerational elasticity of earnings for fathers and sons in the U.S. was estimated to be less than 0.2 among early studies (surveyed in [Becker and Tomes, 1986](#)), ranged between about 0.3 and 0.5 in the studies surveyed in [Solon \(1999\)](#), and is estimated to be around 0.6 or above in more recent studies like [Mazumder \(2005\)](#) and [Gouskova et al. \(2010\)](#).

But neither life-cycle effects as such nor the strategies to address them have yet been evaluated using actual lifetime incomes. In this paper we make use of Swedish data that contain nearly complete income histories of both fathers and sons, allowing us to derive a benchmark estimate and thus to directly expose the bias that results from approximation of lifetime by annual incomes. We test if current empirical practice can reduce this bias and examine how to improve elasticity estimates further.

First, we show that intergenerational elasticity estimates vary substantially with the age at which sons' incomes are observed, confirming that life-cycle effects should be of serious concern. The elasticity is below 0.20 when sons' incomes are measured at age 30 but above 0.40 at age 50, implying a drastically different degree of mobility. Second, life-cycle bias is smallest when incomes are observed around midlife. We thus verify a central implication from Haider and Solon's generalization of the classical errors-in-variables model, which is heavily relied on in the current empirical literature. Third, while there is indeed an age at which the bias is zero, we find that the standard methodology fails to predict this "ideal" age. Small age deviations lead to notable shifts in elasticity estimates, suggesting that current empirical strategies may still be subject to substantial bias. Finally, we examine if modifications of the standard methodology can reduce life-cycle bias further. We present an extension of the generalized model that introduces additional covariates, and find that an explicit consideration of human capital accumulation and low-income episodes suffices to strongly reduce the remaining bias.

Our analysis centers on Haider and Solon's generalization of the textbook errors-in-variables model, which adds an age-dependent slope coefficient to true lifetime incomes but maintains the assumption that the remaining error is uncorrelated with true values. Under this assumption, life-cycle bias is eliminated when the age-dependent slope coefficient converges to one. Unfortunately, our data do not fully support this prediction – at this age, the remaining bias from left-side measurement error alone amounts to about 20 percent of the true elasticity (0.21 vs. 0.27). Conceptually, the prediction's underlying assumption fails to hold because the shape of income profiles differs with parental background even for a given level of lifetime income. Life-cycle bias thus tends to be substantially larger than the generalized model predicts. However, the model is rarely used for formal bias correction, and instead motivates what has become a widely applied rule of thumb in the literature – to measure incomes around mid- instead of early or late age. Our results confirm that this strategy strongly improves intergenerational elasticity

estimates, and illustrates how much bias should be expected to remain in applications.

We then examine if modifications of standard practice can reduce this bias further. We present an extension of the generalized errors-in-variables model in which the relationship between annual and lifetime income is allowed to vary across groups. For example, highly educated individuals are found to deviate substantially from the population-average relationship in their early career. The inclusion of a covariate capturing college education thus considerably improves elasticity estimates that are based on early-age income. We show that the generalized model can be strongly affected by low-income episodes, and that their separate treatment can substantially reduce bias at mid- and old ages. This result helps in turn to explain why life-cycle bias also can be reduced by averaging over multiple income observations on the left-hand side (i.e. for the offspring), a procedure that reduces the influence of low-income episodes.

Our results thus have both positive and negative implications. They corroborate that incomes should be measured in midlife, and that deviating from this rule of thumb will have detrimental consequences. But they also imply that current methods to compensate for incomplete income data are still imperfect, and that mobility estimates are likely less accurate than commonly assumed. Well-established findings from the literature, such that income mobility is lower in the U.S. than in the Nordic countries, are not put into doubt. But attempts to detect more gradual differences, as in recent studies on mobility trends, can be more easily compromised by life-cycle bias. We do find that simple extensions of the generalized errors-in-variables model can strongly reduce the remaining bias. However, some rely on data that may not always be available in practice. We therefore discuss alternative ways for practitioners to make use of our findings.

Life-cycle bias stems generally from the interaction of two factors: heterogeneity in income profiles cannot be fully accounted for, and unobserved idiosyncratic deviations from average profiles correlate with individual and family characteristics. For example, the offspring from poorer families may have higher initial incomes but flatter slopes if credit constraints affect human capital accumulation and job-search behavior in their early career. Such patterns are also of importance for other literatures that depend on measurement of long-run income and income dynamics. Examples include studies on the returns to schooling and the extensive literature that relates measures of stochastic income shocks to consumption or other outcomes.

The next section describes the methodology and identifying assumptions employed in

the early literature. We examine the generalized errors-in-variables model theoretically in section 2 and empirically in section 3. We present and test extensions of that model in section 4, and section 5 concludes.

3.1 The Intergenerational Mobility Literature

The target regression model in intergenerational mobility research is

$$y_{s,i}^* = \beta y_{f,i}^* + \epsilon_i, \quad (3.1)$$

where $y_{s,i}^*$ denotes log lifetime income of the son in family i , $y_{f,i}^*$ log lifetime income of his father, ϵ_i is an error term that is orthogonal to $y_{f,i}^*$, and variables are expressed as deviations from their generational means.³ The coefficient β captures a *statistical relationship* that is commonly referred to as the intergenerational income elasticity.⁴

Approximation of Lifetime Income

As commonly available data sets do not contain complete income histories for two generations, a major challenge is how to approximate lifetime income.⁵ Let y_i be some observed proxy for unobserved log lifetime income of an individual in family i , e.g. a single-year observation, an average of multiple annual income observations, or a more complex estimate based on such annual incomes. Observed values are related to true values by

$$y_{s,i} = y_{s,i}^* + u_{s,i},$$

³We use the terms earnings and income interchangeably (since the issues that arise are similar), and examine fathers and sons since this has been the baseline case in the literature. A growing literature exists on intergenerational mobility in other family dimensions (e.g mothers, daughters or siblings) and in other income concepts (such as household income), for which our conceptual arguments are likewise relevant.

⁴Equations akin to (1) may also appear as *structural relationships* to study *causal mechanisms* of intergenerational transmission. The structural relationship relates typically not to ex-post measures of long-run economic status but to the ex-ante concept of “permanent income”. The two concepts are not always clearly distinguished, and some studies adopt the term “permanent income” even while focusing on the measurement of mobility. Our analysis relates to the statistical relationship, but incomplete measurement of long-run status impedes identification of both types.

⁵Note that the availability of better data would not generally solve the identification problem, since data sets cannot contain complete income histories for contemporary populations.

where $y_{s,i}^*$ is the unobserved true log lifetime income of the son in family i and $u_{s,i}$ is measurement error. Similarly, for the father we observe

$$y_{f,i} = y_{f,i}^* + u_{f,i}.$$

The probability limit of the OLS estimator from a linear regression of y_s on y_f can be decomposed into

$$plim \hat{\beta}_{approx} = \frac{Cov(y_f, y_s)}{Var(y_f)} = \frac{\beta Var(y_f^*) + Cov(y_f^*, u_s) + Cov(y_s^*, u_f) + Cov(u_s, u_f)}{Var(y_f^*) + Var(u_f) + 2 Cov(y_f^*, u_f)}, \quad (3.2)$$

where we used eq. (3.1) to substitute for $y_{s,i}^*$ and applied the covariance restriction $Cov(y_{f,i}^*, \epsilon_i) = 0$. It follows that the estimator can be down- or upward biased and that the covariances between measurement errors and lifetime incomes impact on consistency. The empirical strategies employed in the literature in the last decades can be broadly categorized in terms of changes in identifying assumptions about these covariances.

First Two Waves of Studies

The *first wave* of studies, surveyed in [Becker and Tomes \(1986\)](#), neglected the problem of measurement error in lifetime status. Often just single-year income measures were used as proxies for lifetime income, thereby implicitly assuming that

$$Cov(y_f^*, u_s) = Cov(y_s^*, u_f) = Cov(u_s, u_f) = Cov(y_f^*, u_f) = 0,$$

and

$$Var(u_f) = 0.$$

Classical measurement error in lifetime income violates the latter assumption, so that estimates suffered from large attenuation bias. Estimates of the intergenerational elasticity were therefore too low. This problem was recognized in [Atkinson \(1980\)](#) and then frequently addressed in the *second wave* of studies (surveyed in [Solon 1999](#)). But the assumption remained that measurement errors are random noise, independent of each other and of true lifetime income. That life-cycle variation had to be accounted for was recognized, but it was generally assumed that including age controls in the regression

equation would suffice. The assumptions were therefore

$$Cov(y_f^*, u_s) = Cov(y_s^*, u_f) = Cov(u_s, u_f) = Cov(y_f^*, u_f) = 0,$$

and

$$Var(u_f) \neq 0.$$

If these hold, then the probability limit in eq. (3.2) reduces to

$$plim \hat{\beta}_{approx} = \beta \frac{Var(y_f^*)}{Var(y_f^*) + Var(u_f)}.$$

This is the classical errors-in-variables model; inconsistencies are limited to attenuation bias caused by measurement error in lifetime income of *fathers*. In contrast, measurement error in sons' lifetime income is assumed to not be a source of inconsistency in this model. Researchers typically used averages of multiple income observations for fathers to increase the signal-to-noise ratio, but gave less attention to the measurement of sons' income.

Recent Literature

Recently the focus has shifted towards the importance of non-classical measurement error. An early theoretical discussion can be found in [Jenkins \(1987\)](#). Analyzing a simple model of income formation, he finds that usage of current incomes in eq. (3.1) will bias $\hat{\beta}$ as income growth over the life cycle varies across individuals. He concludes that the direction of this *life-cycle bias* is ambiguous, that it can be large, and that it will not necessarily be smaller if fathers' and sons' incomes are measured at the same age.

[Haider and Solon \(2006\)](#) demonstrate that life-cycle bias can explain the previously noted pattern that intergenerational elasticity estimates increase with the age of sampled sons.⁶ They show that the association between current and lifetime income varies systematically over the life cycle, contrary to a classical errors-in-variables model with measurement error independent of true values. [Böhlmark and Lindquist \(2006\)](#) find strikingly similar patterns in a replication study with Swedish data.

Haider and Solon also note that controlling for the central tendency of income growth

⁶For a summary, see [Solon \(1999\)](#). Age-dependency of elasticity estimates could also arise if the dispersion in transitory income and thus the attenuation bias vary over the life cycle. Such variation has been documented in [Björklund \(1993\)](#) for Sweden, but [Grawe \(2006\)](#) finds that the observed age-dependency can be better explained by the existence of life-cycle bias.

in the population by including age controls in eq. (3.1) will not suffice, as variation around the average growth rate will bias estimates. Vogel (2006) provides an illustration based on the insight that highly educated workers experience steeper-than-average income growth. Since available data tend to cover annual incomes of young sons and old fathers, lifetime incomes of highly educated sons (fathers) will be understated (overstated), which is likely to bias $\hat{\beta}_{approx}$ substantially downwards if educational achievement is correlated within families. Indeed, the probability limit of $\hat{\beta}_{approx}$ can be negative in extreme cases, as our data will confirm. Various refined estimation procedures have been proposed to address such life-cycle bias. We proceed to examine the most popular one in detail.

3.2 Measuring Income at a Certain Age

Haider and Solon (HS) generalize the classical errors-in-variables model to allow for variation in the association between annual and lifetime income over the life cycle, which they document to be substantial. Their underlying intuition is that for two individuals with different income trajectories there will nevertheless exist an age t^* where the difference between individuals' log annual incomes equals the difference between their log (annuitised) lifetime incomes. The generalized model coincides with a classical errors-in-variables model at t^* , suggesting that lifetime incomes should be approximated by annual incomes around this age.

The model is applicable to any analysis that relies on approximation of lifetime income by short-term measures, but we describe it here in the context of the intergenerational mobility literature. Assume that $y_{s,i}^*$ and $y_{f,i}^*$ are unobserved and proxied by $y_{s,it}$ and $y_{f,it}$, log annual incomes at age t . Haider and Solon's generalization of the classical errors-in-variables model is given by the linear projection of $y_{f,it}$ on $y_{f,i}^*$ as

$$y_{f,it} = \lambda_{f,t} y_{f,i}^* + u_{f,it}, \quad (3.3)$$

where $\lambda_{s,t}$ is allowed to vary by age and $u_{s,it}$ is orthogonal to $y_{s,i}^*$ by construction, and similarly the linear projection of $y_{s,it}$ on $y_{s,i}^*$ as

$$y_{s,it} = \lambda_{s,t} y_{s,i}^* + u_{s,it}. \quad (3.4)$$

Under the generalized model, the probability limit of the OLS estimator from a linear

regression of $y_{s,it}$ on $y_{f,it}$ becomes

$$\begin{aligned} \text{plim } \hat{\beta}_t &= \frac{\text{Cov}(y_{s,t}, y_{f,t})}{\text{Var}(y_{f,t})} \\ &= \frac{\beta \lambda_{s,t} \lambda_{f,t} \text{Var}(y_f^*) + \lambda_{f,t} \text{Cov}(y_f^*, u_{s,t}) + \lambda_{s,t} \text{Cov}(y_s^*, u_{f,t}) + \text{Cov}(u_{s,t}, u_{f,t})}{\lambda_{f,t}^2 \text{Var}(y_f^*) + \text{Var}(u_{f,t})}. \end{aligned} \quad (3.5)$$

As HS we first focus on left-side measurement error and assume that $y_{f,i}^*$ is observed (such that $\lambda_{f,t} = 1$ and $u_{f,t} = 0 \forall i$). Then the probability limit in equation (3.5) becomes

$$\text{plim } \hat{\beta}_t = \frac{\text{Cov}(y_{s,t}, y_f^*)}{\text{Var}(y_f^*)} = \beta \lambda_{s,t} + \frac{\text{Corr}(y_f^*, u_{s,t}) \sigma_{u_{s,t}}}{\sigma_{y_f^*}}. \quad (3.6)$$

HS note that under the assumption

$$\text{Corr}(y_f^*, u_{s,t}) = 0, \quad (3.7)$$

left-side measurement error is innocuous for consistency of intergenerational elasticity estimates if lifetime incomes of sons are proxied by annual incomes at an age t^* where $\lambda_{s,t}$ is close to one. Their empirical analysis reveals that for an American cohort born in the early 1930s $\lambda_{s,t}$ is below one for young ages, but close to one around midlife.

The model, often referred to as the *generalized errors-in-variables (GEiV) model*, thus illustrates how life-cycle bias should be expected to vary with age. Apart from providing conceptual insight, this knowledge can be very useful in applications. Researchers often face the problem that long-run outcomes like lifetime income are of theoretical interest, but that available data only contain short snapshots of income. The GEiV model offers a potential remedy since it implies that measurement of income at a certain age might suffice if long-run outcomes are not directly observed. Possible applications are for example the returns to schooling or, as emphasized by HS, the intergenerational mobility literature.

The model has indeed become the standard reference to motivate empirical strategies in the latter, where the implied procedure to measure income around midlife is now common practice.⁷ A variation of the model that relies on the same intuition has been

⁷Among others, in Gouskova et al. (2010) for the US; Björklund et al. (2006, 2009) for Sweden; Nilsen et al. (2012) for Norway; Raaum et al. (2007) for Denmark, Finland, Norway, the UK and the US; Nicoletti and Ermisch (2007) for the UK; Piraino (2007) and Mocetti (2007) for Italy. More examples are covered in the surveys of Björklund and Jäntti (2009) and Black and Devereux (2011).

presented in [Lee and Solon \(2009\)](#).

But as the classical errors-in-variables model, the GEiV model depends critically on assumption (3.7), as also noted by HS. The validity of this assumption has not been examined and much of the current literature tends to assume that following the broad recommendation of measuring incomes in midlife is enough to eliminate or nearly eliminate life-cycle bias in applications. Yet, there are reasons to suspect that assumption (3.7) or similar assumptions might not hold.

Note first that for more than two workers we will generally *not* find an age t^* where annual income is an undistorted approximation of lifetime income. Figure 3.1 illustrates this argument by plotting log income trajectories for workers 1, 2 (as in Figure 1 in HS) and an additional worker 3. At age t_1^* the difference between the annual income trajectories equals the difference in lifetime income for workers 1 and 2, and at age t_2^* for workers 1 and 3. There exists no age where these differences are equal for all three workers at once.⁸ This example illustrates that the parameter $\lambda_{s,t}$ only captures how differences in annual income and differences in lifetime income relate *on average* among *all* workers. Individuals, and groups of individuals, will nevertheless deviate from this average relationship, so that their annual incomes systematically over- or understate their lifetime incomes compared to the rest of the population. A typical example is that highly educated individuals tend to experience steeper income growth over the life cycle, such that their annual incomes understate (overstate) lifetime incomes at young (old) ages relative to individuals with less education.

For intergenerational mobility studies it is crucial that such idiosyncratic deviations might correlate within families or with parental income. For example, sons from poorer families may have higher initial incomes and flatter slopes if credit constraints affect human capital accumulation and job-search behavior in their early career. There are many other reasons to suspect dependency within families: parents can transmit abilities and preferences, or influence their offspring's educational and occupational choices; all of which may affect the shape of income profiles over the life cycle.⁹ The individual association between annual and lifetime income is thus likely to exhibit an intergenerational

⁸This result does not depend on a high degree of complexity in income growth processes, but holds for example also for a simple log-linear income formation model as analyzed in HS (see [Nyblom and Stuhler, 2011](#)).

⁹For example, individual deviations from the average rate of income growth over the life-cycle may be correlated within families, as considered by [Jäntti and Lindahl \(2012\)](#).

correlation itself and cannot be sufficiently captured by a single population parameter like $\lambda_{s,t}$. Assumption (3.7) is then unlikely to hold, the probability limit of $\hat{\beta}_t$ does not equal $\lambda_{s,t}\beta$, and knowledge of the exact life-cycle pattern of $\lambda_{s,t}$ cannot eliminate life-cycle bias.¹⁰ The basic implications of the GEiV model are not impaired by these arguments. It may still represent a large improvement over the classical errors-in-variables model, which we will examine empirically. Our arguments however imply that life-cycle bias remains hard to address and that the search for an “ideal” age to measure income at might not be an entirely satisfying path to follow.

There are various ways to probe our theoretical arguments. One can examine the validity of assumption (3.7) formally by deriving the elements of $u_{s,it}$ for a given income formation model and analyzing its relation to the regressor $y_{f,i}^*$. While it can be shown that $u_{s,it}$ is correlated with $y_{f,i}^*$ even for a simple log-linear income formation model (see Nybom and Stuhler, 2011), such exercises will not be informative on the magnitude of life-cycle bias that should be expected in practice. In the next section we provide instead empirical evidence.

3.3 Empirical Evidence on Life-Cycle Bias

We use Swedish panel data containing nearly life-long income histories to provide direct evidence on the life-cycle bias in estimates of the intergenerational elasticity that are based on annual incomes. We then apply the GEiV model and examine the size of the remaining bias. We evaluate both the rule of thumb to measure incomes at the predicted “ideal age” t^* and the model’s ability to correct elasticity estimates at other ages.

3.3.1 Data Sources and Sample Selection

To the best of our knowledge, Swedish tax registry data offer the longest panel of income data, covering annual incomes across 48 years for a large and representative share of the population. Moreover, a multi-generational register matches children to parents, and census data provide information on schooling and other individual characteristics. All merged together, the data provide a unique possibility to examine life-cycle bias in intergenerational mobility estimation using actual income histories.

¹⁰Corresponding biases arise in the case of right-side measurement error in which unobserved lifetime income of fathers is approximated by annual income or if approximations are made for both fathers and sons, as can be derived from eq. (3.5).

To select our sample, we apply a number of necessary restrictions. As we mainly aim to make a methodological point, we follow the majority of the literature and limit our sample to sons and their biological fathers. To these we merge income data for the years 1960-2007. Since most other income measures are available only from 1968, we use total (pre-tax) income, which is the sum of an individual's labor (and labor-related) earnings, early-age pensions, and net income from business and capital realizations.

Our main analysis is based on sons born 1955-1957. Earlier cohorts could be used, but then we would observe fewer early-career incomes for fathers. Conversely, later cohorts are not included since we want to follow the sons for as long as possible. Moreover, to avoid large differences in the birth year of fathers, we exclude pairs where the father was older than 28 years at the son's birth. Young fathers and first-born sons are thus over-represented in our sample. On other sampling issues we adopt the restrictions applied by HS and [Böhlmark and Lindquist \(2006\)](#).¹¹

Our data come with a couple of drawbacks. To maximize the length of the income histories we use the measure total income, whereas e.g. HS use labor earnings. However, total income is a highly relevant measure of economic status, approximation of lifetime status gives rise to the same methodological challenges, and Böhlmark and Lindquist find that total income and earnings yield similar estimates of life-cycle bias. Further, the use of tax-based data could raise concerns about missing data in the low end of the distribution if individuals have no income to declare. The Swedish system however provides strong incentives to declare some taxable income since doing so is a requirement for eligibility to most social insurance programs. Hence, this concern most likely only applies to a very small share of the population.

Our data also have many advantages. First, they are almost entirely free from attrition. Second, they pertain to all jobs. Third, in contrast to many other studies, our data are not right-censored. Fourth, we use registry data, which is believed to suffer less from reporting errors than survey data. Fifth, and most important, we have annual data from 1960 to 2007, giving us nearly career-long series of income for both sons and their fathers. Overall, we believe that the data are the best available for the purpose of this study.

Our main sample consists of 3504 father-son pairs, with sons' income measured from

¹¹We restrict the sample to fathers and sons who report positive income in at least 10 years. We exclude those who died before age 50, and sons who immigrated to Sweden after age 16 or migrated from Sweden on a long-term basis (at least 10 years).

age 22 to age 50 and fathers' income measured from age 33 to age 65, irrespective of birth years. We express all incomes in 2005 prices, apply an annual discount rate of 2 percent, and divide the sums by the number of non-missing income observations to construct our measures of annuitised lifetime income. Table 3.1 reports descriptive statistics. Rows (2) and (3) show that dispersions in lifetime income are of similar magnitudes for fathers and sons. Rows (4) and (5) provide information on the number of positive income observations. On average there are more than 28 observations for sons, and more than 30 for fathers, with relatively low dispersion in both cases.

3.3.2 Empirical Strategy

To assess the size of life-cycle bias we compare estimates based on annual incomes with a benchmark estimate that is based on lifetime incomes. As in the theoretical discussion we focus on left-side measurement error (i.e., for sons), although we provide brief evidence on life-cycle bias due to right-side (i.e., for fathers) and measurement error on both sides in a later subsection. We do this for two reasons. First, left-side measurement error has until recently been neglected in the literature. Second, life-cycle bias is not confounded by attenuation bias from classical measurement error on the left-hand side, which simplifies the analysis.

We use our measures of log lifetime incomes $y_{f,i}^*$ and $y_{s,i}^*$ to estimate eq. (3.1) by OLS, which yields our benchmark estimate $\hat{\beta}$.¹² We then approximate log lifetime income of sons $y_{s,i}^*$ by log annual income $y_{s,it}$ (left-side measurement error) to estimate

$$y_{s,it} = \beta_t y_{f,i}^* + \epsilon_i,$$

separately for each age t , to obtain a set of estimates $\hat{\beta}_t$. Finally, we estimate eq. (3.4), which provides us with estimates of $\lambda_{s,t}$. None of these estimations include additional controls.

Under the assumptions of the GEiV model, the probability limit of $\hat{\beta}_t$ equals $\lambda_{s,t}\beta$, and using annual income of sons at age t^* where $\lambda_{s,t} = 1$ consistently estimates β . As

¹²Of course, this estimate is not exactly *true* since we still lack some years of income. This does however not affect of our approach to use the estimate as a benchmark. The GEiV model is not restricted to any specific population, and should therefore be applicable to our variant of the Swedish population in which we truncate income profiles at some age. It is nevertheless advantageous that we have long income histories. First, our benchmark estimate will be close to the true value. Second, since the income profiles contain most of the idiosyncratic heterogeneity that leads to life-cycle bias, we expect our estimate of the bias to be representative for a typical application. We provide evidence that our main findings are not sensitive to the exact length of observed income histories in section 3.3.4.

discussed in the previous section, we anticipate $\hat{\beta}_t$ to be biased even after adjustment by $\hat{\lambda}_{s,t}$. The *remaining* life-cycle bias after adjustment by the GEiV model, denoted by $b(t) = \hat{\beta}_t / \hat{\lambda}_{s,t} - \hat{\beta}$, is thus of central interest.¹³ Note that we assume that $\hat{\lambda}_{s,t}$ is known in order to evaluate the model's theoretical capability to adjust for life-cycle bias under favorable conditions. A second (known) source of inconsistency can arise in that the age profile of $\lambda_{s,t}$ will typically not be directly estimable by the researcher.

3.3.3 Empirical Results

We first present estimates of $\lambda_{s,t}$. Figure 3.2 shows that $\hat{\lambda}_{s,t}$ rises over age and crosses one at around age $t^* = 33$. Largely consistent with others, we find that income differences at young (old) age substantially understate (overstate) differences in lifetime income. We note that $\hat{\lambda}_{s,t}$ is close to one only for a short time around age 33, in contrast to the pattern found for older American and Swedish cohorts in HS and [Böhlmark and Lindquist \(2006\)](#) in which $\hat{\lambda}_t$ remains close to one for an extended period through midlife. A general concern is thus that measuring annual income only a few years earlier or later can cause large differences in elasticity estimates.

Our central estimates are presented in Figure 3.3, which plots $\hat{\beta}$ (the benchmark elasticity), $\hat{\beta}_t$ (estimates based on annual income of sons at age t), and $\hat{\beta}_t / \hat{\lambda}_{s,t}$ (estimates at age t adjusted by the GEiV model). Table 3.2 provides additional statistics in the most central age range around t^* . Note that the sample is balanced within (but not across) each age. Zero or missing income observations that are not considered for estimation of $\lambda_{s,t}$ and β_t are not used to estimate β , which is reestimated for each age. The benchmark elasticity thus varies slightly over age. We list our key findings.

First. Our benchmark estimate of the intergenerational elasticity of lifetime income for our Swedish cohort is about 0.27 (see also Table 3.2). This is marginally higher than what most previous studies have found for Sweden, and should be closer to the population parameter due to our nearly complete income profiles.

Second. We confirm that the variation of $\hat{\beta}_t$ over age resembles the pattern in $\hat{\lambda}_{s,t}$, as predicted by the GEiV model. We therefore find that $\hat{\beta}_t$ increases with age and that the life-cycle bias is negative for young and positive for old ages of sons. One of the central predictions of the GEiV model, that current income around midlife is a better proxy for

¹³The arguments of HS relate to the probability limit. In a finite sample we need to consider the distribution of $b(t)$. Reported standard errors for $b(t)$ are based on a Taylor approximation and take the covariance structure of $\hat{\beta}$, $\hat{\beta}_t$, and $\hat{\lambda}_{s,t}$ into account.

lifetime income than income in young or old ages, is thus confirmed.

Third. The magnitude of life-cycle bias stemming from left-side measurement error alone can be striking. For example, the elasticity is below 0.20 when sons' incomes are measured at age 30 but above 0.40 at age 50, thus resulting in drastically different characterizations of the degree of mobility. Analysis based on income below age 26 yields a *negative* elasticity. We therefore find direct evidence on the importance of life-cycle bias in intergenerational mobility estimates that has been discussed in the recent literature.

Fourth. The life-cycle bias is larger than implied by the GEiV model. While the adjustment of estimates according to this model leads on average to sizable improvements, it cannot fully eliminate the bias. This holds true even under the assumption that the central parameters $\lambda_{s,t}$ are directly estimable.

Fifth. The life-cycle bias is not minimized at age t^* , the age at which the current empirical literature aims to measure income, but at an age $t > t^*$. We report a similar pattern for other cohorts in section 3.3.4.

Sixth. The remaining life-cycle bias $\hat{b}(t)$ around age t^* is substantial and significantly different from zero. Table 3.2 shows that $\hat{b}(t)$ is on average around 0.05 over ages 31-35, which corresponds to about 20 percent of our benchmark. Knowledge of age t^* will thus not eliminate life-cycle bias.

Our arguments apply likewise to the extension of the GEiV model presented in Lee and Solon (2009), which has been applied in much of the recent research on mobility trends (see Nybom and Stuhler, 2011).

We briefly compare these empirical results with our theoretical discussion of the determinants of $\hat{b}(t)$. Table 3.3 shows the components of $\hat{b}(t)$ according to eq. (3.6). Variation of $\hat{b}(t)$ over age stems mostly from variation in the residual correlation $Corr(y_f^*, u_{s,t})$, while the ratio $\sigma_{u_{s,t}}/\lambda_{s,t}\sigma_{y_f^*}$ is close to one over most of the life cycle.¹⁴ Seemingly small residual correlations can thus translate into substantive biases. For example, a residual correlation of 0.03 translates into a life-cycle bias of more than 10 percent of the benchmark elasticity.

We provided intuition why the residuals from eq. (3.4) may correlate with parental income in the previous section. For further evidence we examine if the residuals correlate also with various other characteristics, specifically: (i) father's age at birth of his son, (ii) father's education, (iii) son's education, (iv) son's cognitive ability, and (v) son's

¹⁴The previously documented increase in $\lambda_{s,t}$ over age is offset by an increase in $\sigma_{u_{s,t}}$.

country of birth. Table 3.4 describes how each variable is measured and presents the results. Most estimates are significantly different from zero. The residuals correlate particularly strongly with education, implying that the GEiV model cannot capture some of the heterogeneity in income profiles that arises from human capital investment. But the residuals correlate also with other variables, such as ethnic background.¹⁵ The GEiV model should thus not be expected to eliminate life-cycle bias in other literatures, in which interest lies on different explanatory variables. It captures changes in the average association between annual and lifetime income in the population over age, but applications are typically based on comparisons of specific subgroups of the population. The model can then not fully eliminate life-cycle bias since the association between annual and lifetime income varies not only over age, but also over groups defined by parental income, years of schooling, gender, or other characteristics.¹⁶

These results provide guidance for applied research, but some remarks about generalizability are warranted. Life-cycle bias will differ quantitatively across populations. The bias is determined by the degree of systematic differences in income profiles between sons from poor and sons from rich families. This mechanism is likely to vary across cohorts and countries. The question is if observed qualitative patterns over age can nevertheless be generalized. Figure 3.3 demonstrates that income at old age provides a more reliable base for the GEiV model than income at young age. Thus, the relationship between current and lifetime income differs with respect to family background particularly at the beginning of the life cycle. This result is intuitive if one considers potential causal mechanisms of intergenerational transmission. Sons from rich families might acquire more education or face different conditions that particularly affect initial job search (e.g. regarding credit-constraints, family networks, or ex-ante information on labor market characteristics). Such mechanisms are likely to apply to most populations. Although the size of the life-cycle bias is bound to differ across populations, its pattern over age is thus likely to hold more generally. This conclusion is supported by results for other Swedish cohorts as well as direct evidence on the role of human capital investments, both of which

¹⁵The observation that annual incomes in early age tend to understate lifetime incomes for sons born outside Sweden may for example relate to earnings assimilation, the tendency of immigrants to experience lower initial earnings but faster growth than native workers.

¹⁶The observation that the residuals correlate most strongly with education indicates that the GEiV model may perform worse in applications in which education plays a central role. Bhuller et al. (2011) examine life-cycle bias in returns to schooling estimates, and also analyze the applicability of the GEiV model in this context.

will be discussed later on.

Measurement Error on the Right-Hand Side or Both Sides. For conceptual reasons we focused on left-side measurement error, but evidence on the combined effects of life-cycle bias from both sides is also relevant for practitioners. One may ask if we find similar life-cycle effects from the right-hand side, and whether these tend to cancel out or aggravate the effects from left-side measurement error. We now base estimates of β_t on lifetime income of sons and annual income for fathers (right-side measurement error) or annual incomes for both fathers and sons (measurement error on both sides). The probability limit of $\hat{\beta}_t$ is then affected by both attenuation and life-cycle bias. We adjust for both according to the GEIV model. Results are shown in Figures 3.6 and 3.7.¹⁷

Figure 3.6 demonstrates the additional large attenuating effects from right-side measurement error. The remaining life-cycle bias after adjustment by the GEIV model follows a (qualitatively) similar pattern over age as for the case of left-side measurement error. Figure 3.7 shows the remaining life-cycle bias in the case of measurement error on both sides with fathers' and sons' incomes measured at similar ages. The bias is overall larger than for left-side measurement error alone, thus indicating aggravating effects of measurement error on both sides.¹⁸ Importantly, this is also the case when fathers' and sons' incomes are measured at their respective t^* . We again find that the GEIV model is less successful in reducing the bias for early ages and around t^* than for later ages.

3.3.4 Robustness tests

We test various alterations of the estimation procedure to test the sensitivity of our main results. For simplicity we focus again on left-side measurement error only.

Treatment of Outliers in the Income Data. Intergenerational elasticity estimates can be sensitive to how one treats outliers and zero or missing incomes (Couch and Lillard,

¹⁷From equation (3.5), the probability limit of $\hat{\beta}_t$ equals $\theta_{f,t}\beta = (\lambda_{f,t}\sigma_{y_f^*}^2/(\lambda_{f,t}^2\sigma_{y_f^*}^2 + \sigma_{u_{f,t}}^2))\beta$ for right-side and $\lambda_{s,t}\theta_{f,t}\beta$ for both-side measurement error under the assumptions of the GEIV model (assuming $Cov(y_f^*, u_{s,t}) = Cov(y_s^*, u_{f,t}) = Cov(u_{s,t}, u_{f,t}) = 0$). Therefore the remaining life-cycle biases equal $b(\hat{t}) = \hat{\beta}_t/\hat{\theta}_{f,t} - \hat{\beta}$ and $b(\hat{t}) = \hat{\beta}_t/\hat{\lambda}_{s,t}\hat{\theta}_{f,t} - \hat{\beta}$, respectively. For presentational purpose we use only one age subscript t and display combinations of annual income for sons and fathers with equal distances to their respective t^* in Figure 3.7.

¹⁸This holds true if estimates are only adjusted for attenuation bias but not for life-cycle effects according to the GEIV model (see Figure 13 in Nybom and Stuhler, 2011). These results confirm and substantiate the theoretical predictions of Jenkins (1987) that measuring fathers' and sons' income at similar ages might not necessarily reduce life-cycle bias.

1998; Dahl and DeLeire, 2008). We test the robustness of our results by (i) balancing the sample such that only sons with positive income in all ages 31-35 are included, (ii) bottom-coding very low non-missing incomes, and (iii) top-coding very high incomes. We compare the life-cycle bias at ages 31-35 for each of these samples (summarized in Table 3.5) with the results for our main sample (Table 3.2). Estimates of the remaining life-cycle bias are on average a third lower for the balanced than for our main sample (across ages 31-35), but still correspond to more than 10 percent of the benchmark elasticity. Decreases in both the residual correlation and residual variance contribute to this drop. Bottom-coding increases the bias slightly, perhaps since observations with zero income are now always included. Finally, results for the top-coded sample are very similar to those for the main sample, implying low sensitivity to the exact measurement of high incomes. Zero and low incomes thus seem influential for the size of life-cycle bias, but it is not obvious what the right sampling choice would be. We will come back to this question in the next section.

Length of Observed Income Profiles. Although our data are the best available for our purpose, it might be a concern that our measures of lifetime income are only based on *almost* complete income histories. We thus test if our findings are sensitive to the exclusion of short spans of income data. Age profiles of the life-cycle bias before and after adjustment by $\hat{\lambda}_{s,t}$ for various such tests are shown in Figure 3.4.¹⁹ Changes in the fathers' age span have little effect on the life-cycle bias. Changes in the sons' age span cause noticeable shifts, although the pattern over age remains stable. This is not unexpected since such changes are likely to alter both $\sigma_{y_s^*}$ and $\lambda_{s,t}$ slightly. The exact relation between the life-cycle bias and age therefore depends on the observed age span, but the major facts remain stable: the remaining life-cycle bias after adjustment by $\hat{\lambda}_{s,t}$ can be large and tends to be negative for young ages and around t^* .

Cohort and Population Differences. We repeat our analysis for two other cohort groups (sons born 1952-54 or 1958-60), which will also illustrate if the magnitude of life-cycle bias varies across populations. To separate cohort from age differences, we limit income profiles to the longest age span observed in all three samples (ages 22-47 for sons

¹⁹In our working paper (Nyblom and Stuhler, 2011) we also consider younger cohorts — sons born 1958-60 — to study the influence of early-age income data of fathers, and older cohorts — born 1952-54 — to study the influence of late-age data of sons. The results support the findings reported here.

and 36-65 for fathers). Table 3.6 presents the most central results around age t^* . The 1958-60 cohort has a benchmark elasticity $\hat{\beta}$ that is similar to our main cohort but a slightly larger remaining life-cycle bias $b(\hat{t})$. For the 1952-54 cohort both $\hat{\beta}$ and $b(\hat{t})$ are substantially lower. Figure 3.5 plots estimates of β_t for all three samples over the full age range. While the overall patterns are relatively similar, the differences between elasticity estimates at each age are quite volatile. These differences – substantial even for large random samples and a fixed sampling procedure across cohorts within Sweden – confirm that the degree of bias in elasticity estimates may differ substantially across populations even if incomes are measured at the same age.

3.4 Extensions

We proceed to examine if alterations of the generalized error-in-variables model and standard estimation procedures can reduce life-cycle bias further. We first offer an extension of the generalized model that introduces additional covariates in equation (3.4). Whether averaging over multiple annual income observations also on the left-hand side can reduce bias is addressed in the following subsection.

3.4.1 Extending the Generalized Errors-in-Variables Model

The model presented by Haider and Solon (2006) captures how differences in annual and lifetime incomes relate on average in the population of interest. We showed that knowledge of the average relationship may not be sufficient to eliminate life-cycle bias in applications, as idiosyncratic deviations from this average relate systematically to parental background and other variables. It appears useful to condition this relationship on additional covariates that may capture such heterogeneity.

We thus extend the generalized model by introducing heterogeneous intercepts in equation (3.4) that vary with covariates $x_{1,it}, \dots, x_{\bar{k},it}$, such that

$$y_{s,it} = \lambda_{s,t} y_{s,i}^* + \sum_{k=1}^{\bar{k}} \mu_{k,t} x_{k,it} + u_{s,it}. \quad (3.8)$$

The coefficients $\mu_{1,t}, \dots, \mu_{\bar{k},t}$ capture whether beyond the population-average relationship, annual incomes systematically under- or overstate lifetime incomes for certain groups. Maintaining the assumption $Cov(u_{s,it}, y_{s,i}^*) = 0$, the slope coefficient in a linear regression

of log annual income $y_{s,it}$ on log lifetime income $y_{f,i}^*$ equals

$$\beta_t = \frac{Cov(y_{s,t}, y_f^*)}{Var(y_f^*)} = \beta \lambda_{s,t} + \sum_{k=1}^{\bar{k}} \mu_{k,t} \gamma_{k,t}, \quad (3.9)$$

where

$$\gamma_{k,t} = \frac{Cov(x_{k,t}, y_f^*)}{Var(y_f^*)} \quad \forall k \in \{1, \dots, \bar{k}\}$$

is the slope coefficient from a linear regression of covariate $x_{k,t}$ on father's log lifetime income y_f^* . The true intergenerational elasticity is then equal to

$$\beta = \frac{Cov(y_s^*, y_f^*)}{Var(y_f^*)} = \frac{\beta_t - \sum_{k=1}^{\bar{k}} \mu_{k,t} \gamma_{k,t}}{\lambda_t}. \quad (3.10)$$

We will examine if such extension of the generalized model performs better, both in predicting the “ideal” age to measure incomes at and in minimizing the bias when using income observations beyond the “ideal” age or age range.

In the previous section we found the bias to be particularly large at young age, and noted that differential human-capital investment is a potential explanation. We thus first consider information on educational attainment. We include a single covariate $e_{s,i}$ that equals one if a son has attended university or college and zero otherwise in equation (3.8), such that the annual elasticity equals $\beta_t = \beta \lambda_{s,t} + \mu_{e,t} \gamma_{e,t}$ if $Cov(u_{s,it}, y_{s,i}^*) = 0$ holds. Estimating all coefficients by OLS, the probability limit of $(\hat{\beta}_t - \hat{\mu}_{e,t} \hat{\gamma}_{e,t}) / \hat{\lambda}_{s,t}$ then equals the true elasticity β .

Panel (A) of Figure 3.8 shows that the life-cycle profile of λ_s is similar for the standard GEiV (solid line) and our extended model (dashed line). Panel (B) presents the life-cycle profile of μ_e (dashed line), illustrating that the income profile of highly educated individuals deviate strongly from the average relationship that is captured in λ_s : their annual incomes understate lifetime incomes strongly in early and overstate them in old ages, even after the average relationship between annual and lifetime income is taken into account.

College attendance is highly correlated with parental income (in our data $\hat{\gamma}_{e,t} \approx 0.23 \forall t$), such that the explicit consideration of these deviations may improve intergenerational elasticity estimates. Figure 3.9 compares the remaining life-cycle bias in annual elasticity estimates after adjustments based on the standard GEiV (solid line) or our extended

model (dashed line). The extension strongly reduces life-cycle bias in early age; estimates are in the vicinity of the true elasticity from the mid-twenties, in contrast to the standard model. This is not surprising: early-age income is a poor signal of lifetime income particularly for college graduates, such that the approximation of lifetime by annual income differences is strongly distorted when education is not taken into account. However, controlling for education reduces the bias only mildly around t^* and the intergenerational elasticity is still underestimated at mid- and old ages.

Post-secondary education is one of various sources for low-income episodes, which transformed to logs can be highly influential in the estimation of $\lambda_{s,t}$.²⁰ It may be beneficial to treat those observations separately, in particular as from about age 30 they are exceedingly rare among sons with high lifetime income. Adding a second covariate that equals one if $y_{s,t}$ is below the first percentile in equation (3.8) strongly reduces estimates of $\lambda_{s,t}$ (Figure 3.8, panel (A), short-dashed line). The effect of the college dummy is thus predominantly captured by $\mu_{k,t}$, while the low-income dummy also has a large influence on estimates of $\lambda_{s,t}$. Figure 3.9 shows that this model performs much better throughout midlife. While addition of this covariate alone fails to significantly improve estimates in early ages, the adjusted elasticity based on annual incomes is otherwise surprisingly accurate.

But how can practitioners make use of these findings? Application of the above procedure is straightforward if the data allow for direct estimation of equation (3.8).²¹ If the relationship between annual and lifetime incomes cannot be directly estimated then external evidence can potentially be combined with own estimates. For example, external estimates of $\lambda_{s,t}$ and $\mu_{e,t}$ could be combined with own estimates of $\gamma_{e,t}$ if the intergenerational data set includes educational information. As noted by Haider and Solon, importing estimates from other data sets can be problematic if the central relationships differ across populations. But while income-education associations vary considerably across countries we may expect similarities in the broader patterns (e.g. education decreasing income at early but increasing at late age). Our finding that exceptionally low incomes should be treated separately can also be adopted even when

²⁰Other potential sources include episodes of unemployment or long-term sick leave, of voluntary leisure or non-market work, and time spent abroad.

²¹For example, the intergenerational data may include educational attainment, annual and lifetime incomes for one generation but not the other. Alternatively, an external data source that only covers one generation of the population of interest may be exploited for this purpose.

equation (3.8) cannot be directly estimated. One can discard low-income episodes in the estimation of β_t and for bias-correction use estimates of $\lambda_{s,t}$ from samples in which those observations have been likewise discarded. We find that it makes little difference if low-income episodes are simply dropped in both the estimation of (3.8) and (3.9) or included but treated separately, as described above.

We conclude that simple extensions of Haider and Solon’s model can substantially reduce life-cycle bias in applications: the average absolute deviation in adjusted annual estimates across ages 25-45 corresponds to 23.6 percent of the true elasticity for the standard model, but falls to 13.8 percent with education or 9.0 percent with education and low-income controls. Moreover, the remaining deviations are centered around instead of below zero in the extended model. The improvements are most striking in young age, at which differences in annual incomes signal differences in lifetime income poorly and the explicit consideration of human capital investments becomes essential. Note however that we cannot reduce bias much further by using more detailed information (on finer educational classes, cognitive ability, ethnicity, or location).²² This observation implies that the income trajectories of sons from high- and low-income fathers remain different even when many individual characteristics can be controlled for. However, the consequences are much reduced, and seem quantitatively important mainly in early and late life. In addition, only few practitioners can do these bias corrections themselves, and our results here can then probably only provide partial improvements and prevent the most extreme type of biases. Life-cycle bias will thus often remain a concern in applications.

3.4.2 Multi-Year Averages of Current Income

The importance of dealing with transitory noise in short-run income measures on the right-hand side, for example by using multi-year averages, is well recognized in the literature (see Mazumder, 2005). But some recent studies that reference to the GEiV model (see footnote 7) average also over multiple income observations on the left-hand side (e.g. for sons). Yet, the theoretical motivation for doing so is not clear. One rationale could be that researchers do not know the exact age at which $\lambda_{s,t}$ equals one. Our finding that life-cycle bias can be substantial even at this age raises the question if

²²Variants of equation (3.8) that allow for the slope parameter $\lambda_{s,t}$ to differ across groups did not perform better than the simpler specification with heterogeneous intercepts, and are thus not presented here.

and how averaging can help to reduce the bias.

We therefore estimate β_t using the logs of three-, five- and seven-year averages of son's income. These averages are also used to estimate $\lambda_{s,t}$ and the remaining life-cycle bias after adjustment by $\hat{\lambda}_{s,t}$. Figure 3.10 presents its size for averages that are centered around different ages. The remaining life-cycle bias falls in the number of income observations but is not eliminated. With seven-year averages the true elasticity is underestimated by about 0.03 at ages 31-35 compared to about 0.05 using one-year measures. The standard deviation of the residuals $\hat{\sigma}_{u,s,t}$, which is a central component of the bias, decreases by about a third as we move from one- to seven-year measures, and diminishes the estimated bias proportionally. The residual correlation falls only slightly and estimates of $\lambda_{s,t}$ are marginally lowered up until about age 40. As of the log transformation, averaging reduces the influence of episodes with very low incomes, which in the previous section were found to contribute to life-cycle bias in the GEiV framework.

In addition, estimates based on annual measures may suffer from strong year-to-year variability (see Figure 3.3). Reducing this variability is a second motive for averaging over multiple income observations on both sides. Our results thus provide two separate arguments in support of averaging over income observations also on the left-hand side, when possible. Note however that these results pertain to using log of multi-year averages, not to multi-year averages of log annual incomes. As noted by Haider and Solon (2006), estimates based on the latter are algebraically equivalent to the simple average of the single-year estimates, and will thus only smooth estimates.

3.5 Conclusions

Using snapshots of income over shorter periods in the estimation of intergenerational income elasticities causes a so-called life-cycle bias if the snapshots cannot mimic lifetime outcomes (Jenkins, 1987). We use nearly career-long income data of fathers and their sons, allowing us to estimate a benchmark elasticity and to directly expose the large magnitude of this bias in practice. We confirm that Haider and Solon's (2006) generalization of the classical errors-in-variables model and their widely adopted suggestion to measure incomes around mid-age can strongly improve elasticity estimates. However, we also show that the failure of another errors-in-variables assumption prevents correct prediction of the ideal age of measurement and thus full elimination of life-cycle effects.

The bias that persists in our Swedish data even after application of the generalized model is strongly negative when using annual income below age thirty and remains negative up until the early forties. Estimates understate the true elasticity substantially also when income is measured around the “ideal” age as predicted by the generalized model.

Comparisons of intergenerational mobility estimates across countries, groups or cohorts may thus be of limited reliability if based on short-run income data.²³ Still, some of the major conclusions from cross-country studies are not put into question. For example, the findings that income mobility is much lower than found by the early literature, and that mobility differs strongly across countries (e.g. being lower in the U.S. than in the Nordic countries and Canada), would be robust even to sizable revisions in the underlying estimates. It might however be necessary to revisit those conclusions that are based on more marginal differences. Studies on mobility trends are potentially affected since even moderate life-cycle biases may be sufficient to mask gradual changes of mobility over time. Comparisons across subgroups of a population can be compromised when the age pattern in income profiles differs, which may for example be the case when groups are classified by education, sex or immigration status.

These results are mostly negative, but our analysis also points to potential improvements. We find evidence that incomes at later ages (e.g. age 40-50) provide a more reliable base for application of the GEiV model. Moreover, the bias can be reduced by averaging over multiple income observations from midlife (if available) for *both* fathers and sons. Using logs of multi-year averages also for sons may counteract the disproportionate influence of occasional low-income episodes, and has the added value of reducing year-to-year volatility from annual measures.

These simple suggestions lead to modest bias reductions, but we also propose and test a simple extension of Haider and Solon’s generalized errors-in-variables model that may improve elasticity estimates more substantially. The standard model captures how differences in annual and lifetime incomes relate on average in the population, but knowledge of the average relationship is not sufficient if idiosyncratic deviations relate systematically to parental background and other factors. We show how additional

²³One might hope that the bias is of similar magnitude across populations, such that the validity of comparative studies is not affected. Cross-country comparisons would for example be reliable if both the dispersion and the intergenerational correlation in the shape of income profiles is of the same magnitude in each country. But since the intergenerational correlation in income levels varies across countries we suspect that it also differs in other dimensions of income profiles. Our finding that the life-cycle bias varies even across Swedish cohorts born in the same decade supports this conclusion.

covariates can be incorporated into the model to capture some of those deviations. We find that the large bias in early age can be reduced considerably by conditioning the average relationship on education. This suggests that human capital investments are one reason why early-career differences in income predict lifetime differences so poorly, and that the explicit consideration of such investments can strongly increase the signaling value of short-run income differentials. Bias at later ages can be substantially reduced by separate treatment of low-income episodes. Transformed to logs they are highly influential in the estimation of the central parameter of the generalized model, and may generate large life-cycle biases depending on how they relate to parental income. However, not all of those improvements can be realized in practice, as some depend on the availability of (external) evidence on the relationship between lifetime income and individual characteristics, such as education.

Further refinements of empirical practice with restricted use of income observations around a specific age can thus improve upon previous estimates, but will typically not eliminate life-cycle bias. Development of a more structured approach that aims to capitalize on all available income data seems desirable. Future research could in particular benefit from a more comprehensive exploitation of partially observed income growth patterns. Intergenerational mobility estimates are often based on multiple income observations per individual, but researchers tend to disregard the idiosyncratic income growth across these observations. Such partially observed growth patterns are determined by both observable and unobservable characteristics of the individual and may hence contain more information on lifetime income than what current income levels and observable characteristics can provide.

Our results add to a general conclusion that can be drawn from the intergenerational mobility literature: addressing heterogeneity in income profiles is an important, difficult and recurrently underestimated task. The widespread practice of measuring annual income at a certain age as a surrogate for unobserved lifetime income is still prone to life-cycle bias, since the most appropriate age for measurement is hard to predict and since estimates can be sensitive to small age changes. These issues are potentially important for other literatures that rely on measurement of long-run income or income dynamics.

Tables and Figures

Table 3.1: Summary Statistics by Birth Year of Sons

	All	1955	1956	1957
Father's age at birth of son	24.68 (2.53)	24.66 (2.51)	24.77 (2.50)	24.62 (2.58)
log lifetime income (sons)	11.97 (0.43)	11.98 (0.42)	11.98 (0.42)	11.95 (0.44)
log lifetime income (fathers)	11.72 (0.42)	11.73 (0.44)	11.72 (0.43)	11.72 (0.40)
# of pos. income obs. (sons)	28.52 (1.86)	28.57 (1.71)	28.56 (1.74)	28.43 (2.11)
# of pos. income obs. (fathers)	30.32 (3.76)	29.99 (4.13)	30.36 (3.62)	30.59 (3.48)
Father-son pairs (N)	3504	1167	1173	1164

Notes: The table reports means with standard deviations within parentheses.

Table 3.2: OLS Estimates of Elasticities and Life-Cycle Bias

t=Age	$\hat{\lambda}_{s,t}$	$\hat{\beta}$	$\hat{\beta}_t$	$\hat{\beta}_t/\hat{\lambda}_{s,t}$	$b(\hat{t})$	$b(\hat{t})$ in %	N
31	0.897 (0.019)	0.266 (0.016)	0.191 (0.023)	0.213 (0.029)	-0.053 (0.021)	19.8	3478
32	0.909 (0.018)	0.267 (0.016)	0.246 (0.023)	0.271 (0.028)	0.003 (0.021)	1.3	3476
33	0.982 (0.020)	0.267 (0.016)	0.203 (0.025)	0.207 (0.031)	-0.061 (0.024)	22.7	3479
34	1.039 (0.019)	0.256 (0.016)	0.212 (0.025)	0.204 (0.031)	-0.051 (0.023)	20.1	3469
35	1.114 (0.021)	0.261 (0.016)	0.234 (0.027)	0.210 (0.029)	-0.052 (0.022)	19.7	3460

Notes: Cohort group 1955-1957, left-side measurement error only. The sample and thus the benchmark estimate $\hat{\beta}$ are allowed to vary by age due to partially missing data. Standard errors in parentheses, which for $\hat{\beta}_t/\hat{\lambda}_{s,t}$ and $b(\hat{t})$ are based on Taylor approximations that take the covariance structure of $\hat{\lambda}_{s,t}$, $\hat{\beta}$, and $\hat{\beta}_t$ into account. Column (7) displays $b(\hat{t})$ in percent of the benchmark estimate $\hat{\beta}$.

Table 3.3: Decomposition of Life-Cycle Bias

t=Age	$\hat{b}(t)$	$Corr(y_f^*, \hat{u}_{s,t})$	$\hat{\sigma}_{u_{s,t}}$	$\hat{\sigma}_{y_f^*}$	$\hat{\sigma}_{u_{s,t}} / \hat{\lambda}_{s,t} \hat{\sigma}_{y_f^*}$
31	-0.053	-0.044	0.455	0.424	1.198
32	0.003	0.003	0.431	0.423	1.123
33	-0.061	-0.052	0.485	0.422	1.169
34	-0.051	-0.050	0.452	0.422	1.031
35	-0.052	-0.049	0.494	0.422	1.050

Notes: The table displays the remaining bias, $\hat{b}(t)$, together with its associated components. Results are for cohort group 1955-1957, left-side measurement error only.

Table 3.4: Correlations Between Residuals and Characteristics

	Age Interval of Sons				
	26-30	31-35	36-40	41-45	46-50
Father's log lifetime income	-0.057*	-0.050*	-0.063*	-0.020	-0.007
Father's age at birth of son	-0.054*	0.014	0.045*	0.017	-0.006
Father's education	-0.158*	-0.061*	-0.045*	0.035	0.028
Son's education	-0.278*	-0.112*	-0.002	0.085*	0.088*
Son's cognitive ability	-0.108*	-0.073*	-0.050*	0.022	-0.004
Son's country of birth	-0.040*	-0.026	-0.002	-0.032	0.028

Table reports correlations between characteristics listed in the first column and sons' income residuals (as average in each five-year year age interval) from eq. (3.4) for cohort group 1955-1957. The education variables are years of education measured at about age 35, "Son's country of birth" is an indicator for being born outside Sweden, and "Son's cognitive ability" is a standardized cognitive ability measure from the military enlistment cognitive test at age 18. Star superscripts indicate correlations with p-value<0.05.

Table 3.5: Summary of Robustness Tests

t=Age	Balanced Sample			Bottom-Coded Incomes			Top-Coded Incomes		
	$\hat{\beta}$	$\hat{\beta}_t$	$\hat{b}(t)$	$\hat{\beta}$	$\hat{\beta}_t$	$\hat{b}(t)$	$\hat{\beta}$	$\hat{\beta}_t$	$\hat{b}(t)$
31	0.257 (0.015)	0.184 (0.021)	-0.033 (0.021)	0.271 (0.016)	0.205 (0.023)	-0.046 (0.020)	0.251 (0.016)	0.191 (0.024)	-0.050 (0.020)
32	0.257 (0.015)	0.227 (0.020)	0.014 (0.021)	0.271 (0.016)	0.245 (0.023)	-0.012 (0.020)	0.252 (0.016)	0.246 (0.023)	0.005 (0.019)
33	0.257 (0.015)	0.185 (0.023)	-0.053 (0.025)	0.270 (0.016)	0.201 (0.024)	-0.069 (0.022)	0.252 (0.016)	0.203 (0.025)	-0.056 (0.023)
34	0.257 (0.015)	0.219 (0.022)	-0.029 (0.021)	0.270 (0.016)	0.247 (0.025)	-0.044 (0.020)	0.240 (0.015)	0.213 (0.025)	-0.047 (0.022)
35	0.257 (0.015)	0.239 (0.024)	-0.027 (0.020)	0.270 (0.016)	0.250 (0.026)	-0.053 (0.019)	0.246 (0.015)	0.234 (0.027)	-0.048 (0.020)

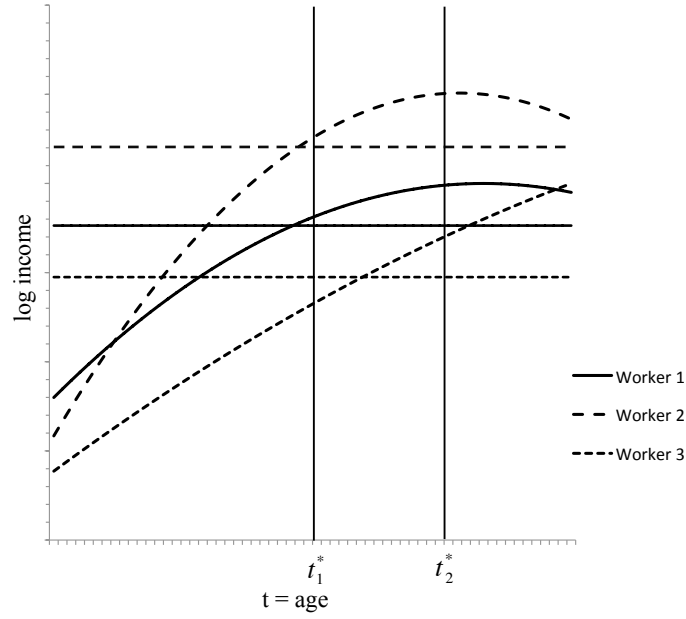
Notes: Cohort group 1955-1957, left-side measurement error only. Standard errors in parentheses. The sample in columns (1)-(3) is balanced across ages, hence excluding individuals who have zero or missing incomes at any age 31-35. The sample in columns (4)-(6) is with low non-missing incomes bottom-coded as 10 000 SEK. The sample in columns (7)-(9) is with high incomes top-coded as 2 000 000 SEK.

Table 3.6: Summary of Cohort Differences, Averages over Ages 31-35

Cohort Group	$\hat{\lambda}_{s,t}$	$\hat{\beta}$	$\hat{\beta}_t$	$\hat{\beta}_t/\hat{\lambda}_{s,t}$	$\hat{b}(t)$	$\hat{b}(t)$ in %	N
1958-60	1.071 (0.022)	0.274 (0.016)	0.235 (0.028)	0.220 (0.032)	-0.054 (0.026)	19.9	3427
1955-57	1.066 (0.020)	0.246 (0.015)	0.216 (0.024)	0.204 (0.028)	-0.042 (0.020)	17.2	3444
1952-54	1.059 (0.021)	0.206 (0.015)	0.190 (0.024)	0.179 (0.027)	-0.027 (0.019)	12.8	3160

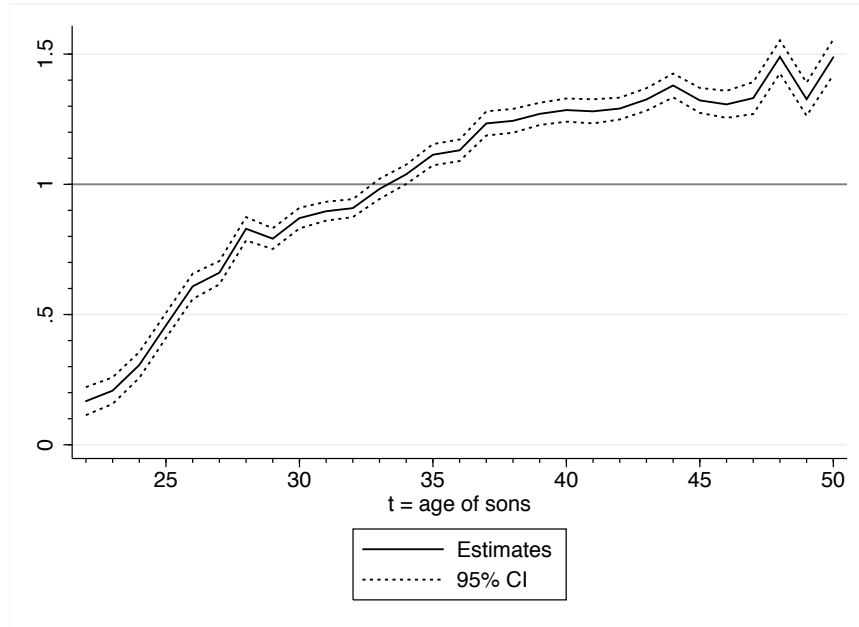
Notes: Left-side measurement error only. Table displays averages of estimates and standard errors (in parentheses) across ages 31-35. $\hat{b}(t)$ is significantly different from zero (p-value<0.05) at three ages (out of five) for 1958-60, at four ages for 1955-57, and at two ages for 1952-54. For all cohort groups, lifetime income is restricted to be measured over identical ages: 22-47 for sons, and 36-65 for fathers. Column (7) displays $\hat{b}(t)$ in percent of our benchmark estimate $\hat{\beta}$ (as average over the age interval).

Figure 3.1: Illustrative Example of Log Annual Income Trajectories



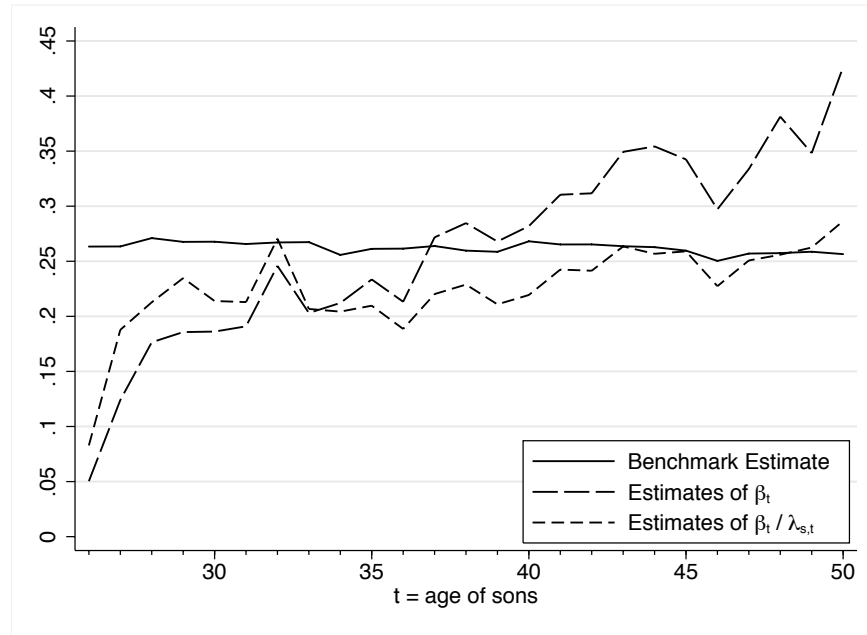
Notes: Illustrative Example. For each worker, the upward-sloping line depicts log annual income by age, the horizontal line depicts log annuitised lifetime income.

Figure 3.2: OLS Estimates of $\lambda_{s,t}$



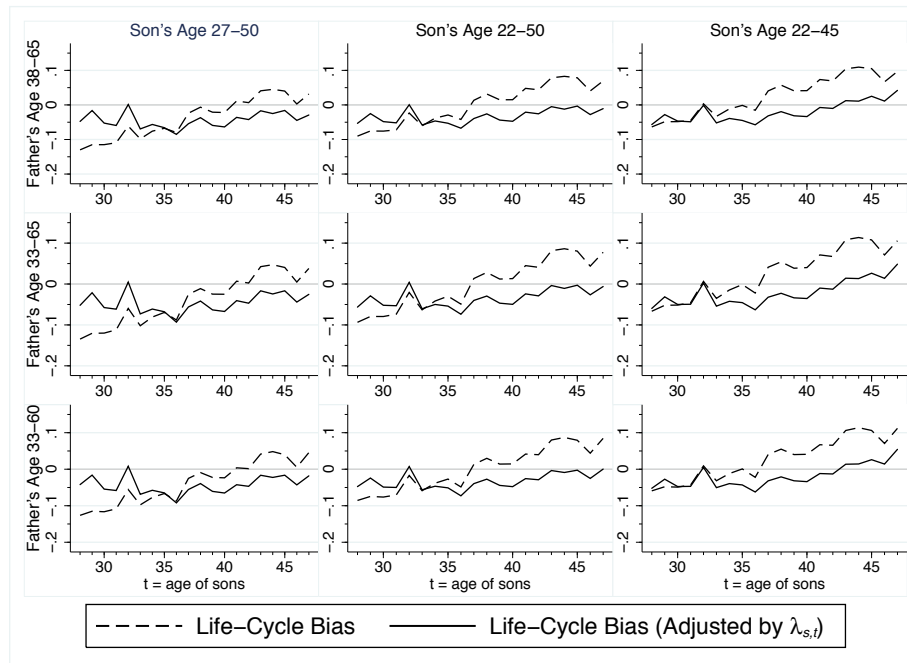
Notes: The figure shows estimates of $\lambda_{s,t}$ by sons' age for cohorts 1955-57. $\lambda_{s,t}$ is the regression coefficient in a regression of son's log annual income on son's log lifetime income, see eq. (3.4).

Figure 3.3: OLS Estimates of Elasticities and Life-Cycle Bias



Notes: The figure shows the benchmark estimate of the intergenerational elasticity together with the unadjusted and adjusted (by the GEIV model) estimates based on sons' annual income. The estimates are for cohort 1955-57, left-side measurement error only.

Figure 3.4: Estimates of Life-Cycle Bias for Different Age Spans (Cohort 1955-57)



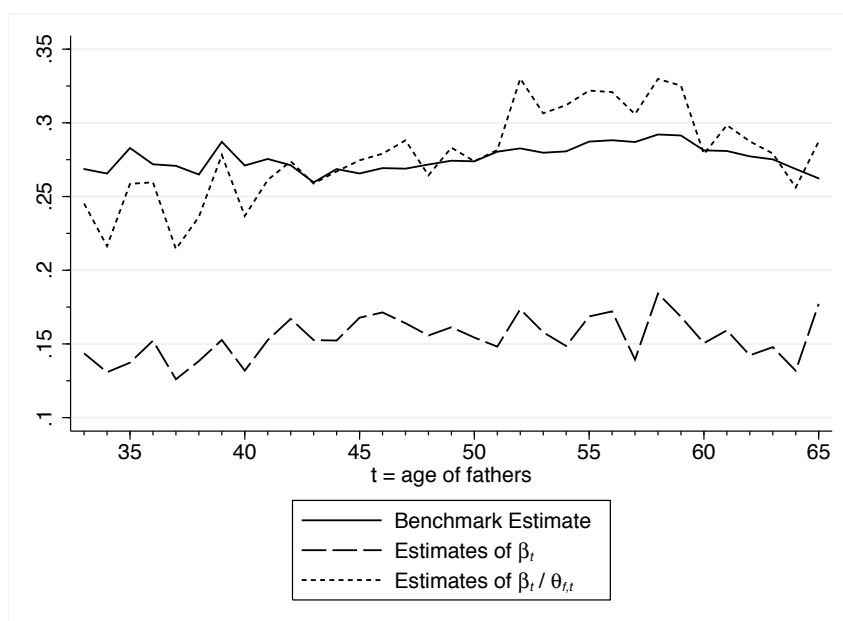
Notes: Left-side measurement error only. The age span of observed incomes of sons (fathers) varies along the horizontal (vertical) dimension.

Figure 3.5: OLS Estimates of Elasticities for Various Cohorts



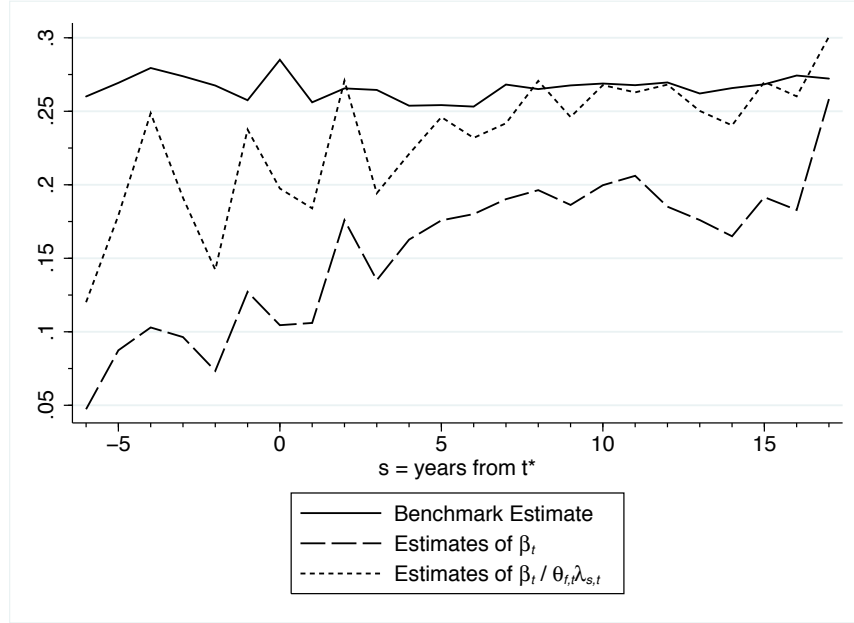
Notes: Left-side measurement error only.

Figure 3.6: OLS Estimates of Elasticities with Right-Side Measurement Error



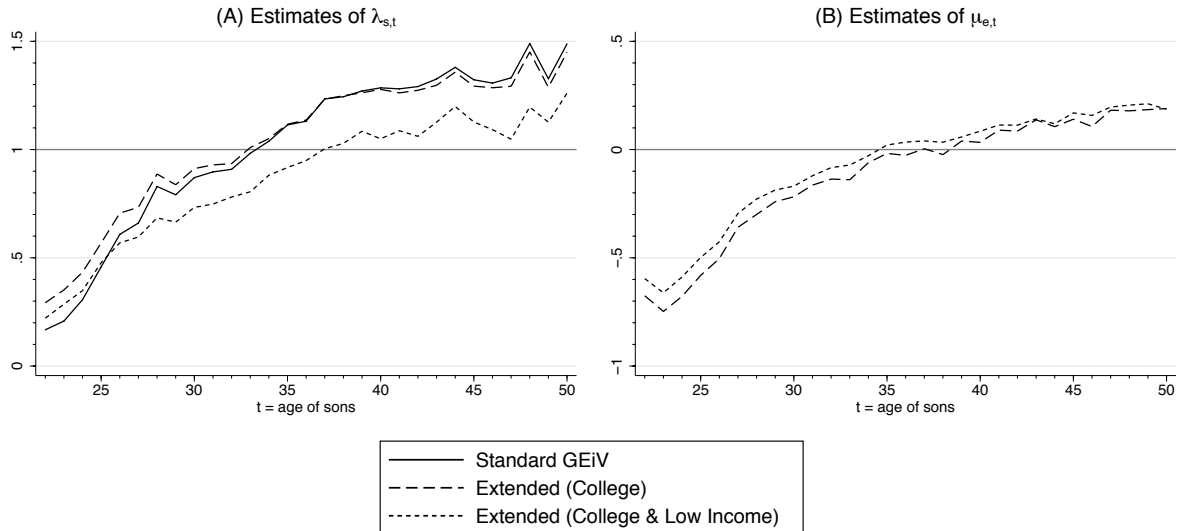
Notes: Cohort 1955-57, right-side measurement error only.

Figure 3.7: OLS Estimates of Elasticities with Both-Side Measurement Error



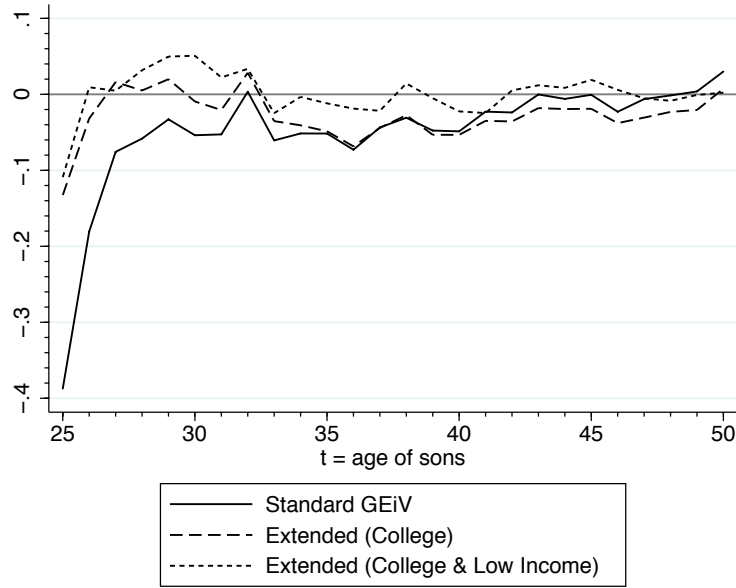
Notes: Cohort 1955-57, measurement error on both sides. For simple presentation we only display results for annual incomes at the same distance from t^* for sons and fathers. At $s=0$ both are measured at their respective t^* , at $s=5$ both are measured five years after t^* , etc.

Figure 3.8: OLS Estimates of Coefficients in Standard and Extended GEiV Model



Notes: Panel(A) shows estimates of $\lambda_{s,t}$ by sons' age for for the standard GEiV model (Haider and Solon, 2006) and extended variants with additional covariates (see main text). Panel (B) shows estimates of $\mu_{e,t}$ for the extended specifications. Cohort 1955-57, left-side measurement error only.

Figure 3.9: Remaining Life-Cycle Bias in Standard and Extended GEiV Model



Notes: The figure shows the difference between the benchmark estimate of the intergenerational elasticity and adjusted estimates based on sons' annual income for the standard GEiV model (Haider and Solon, 2006) and extended variants with additional covariates (see main text). Cohort 1955-57, left-side measurement error only.

Figure 3.10: Remaining Life-Cycle Bias in GEiV Model with Multi-Year Averages



Notes: The figure shows the difference between the benchmark estimate of the intergenerational elasticity and GEiV-adjusted estimates based on the log of three-year, five-year or seven-year averages of sons' annual income (see main text). Cohort 1955-57, left-side measurement error only.

Chapter 4

Mobility Across Multiple

Generations:

The Iterated Regression Fallacy

A vast empirical literature has estimated the degree of intergenerational persistence in socio-economic characteristics between parents and their children. There exists however much less evidence on the degree of *long-run* mobility across *multiple* generations, such as between grandparents and their grandchildren. In its absence we rely on extrapolations from parent-child correlations. For example, [Hertz \(2006\)](#) reports an intergenerational income elasticity of 0.47 for the United States and proceeds to note:

“To understand what these statistics mean, consider a rich and a poor family in the United States [...] and ask how much of the difference in the parents’ incomes would be transmitted, on average, to their grandchildren. In the United States this would be $(0.47)^2$ or 22 percent;”

This procedure – extrapolation by exponentiation – shapes our interpretation of the intergenerational evidence as it is common in policy reports, standard textbooks ([Borjas, 2009b](#)) and specialised survey articles ([Piketty, 2000](#)).

This interpretation matters, as the persistence of economic status across generations is a central aspect in sociological, economic and political theory. [Erikson and Goldthorpe \(1992\)](#) note that competing political theories contain strong and opposing hypotheses about its extent in industrialised societies. [Piketty \(2000\)](#) observes that conflicting views feature also prominently in economic writings. Measuring multigenerational persistence

may help to discriminate between competing schools of thoughts, but it also matters on a practical level. We may for example wonder if specific social policies mask inequalities between families only temporarily or if they have lasting effects on their relative fortunes.

But conflicting views about the degree of long-run mobility persist because we lack direct empirical evidence. Our knowledge about intergenerational mobility on the other hand has advanced greatly in the last two decades. The finding that income mobility is much lower than previously believed, and particularly low in countries with high levels of cross-sectional inequality in which it is more consequential (e.g. [Corak, 2013](#)), has been received with some concern. But the standard extrapolation procedure provides ammunition for a contrarian standpoint that disputes the significance of those findings, as it implies high long-run mobility even when parent-child mobility is low (see for example [Mankiw, 2006](#)).

Its prevalence may seem puzzling, as [Hodge \(1966\)](#) already notes that mobility may not be well described by a first-order Markov process¹. It can perhaps be explained by three factors. First, no comprehensive study exists on how intergenerational and multigenerational mobility *should* relate. In its absence, the iteration of intergenerational measures may be a pragmatic response. Second, this procedure appears prominently in influential studies in the literature. Section IV of [Becker and Tomes \(1979\)](#) draw attention to special *theoretical* cases that vindicate the iteration of intergenerational measures, quoting the old proverb “*from shirtsleeves to shirtsleeves in four generations*” to illustrate its implications. [Becker and Tomes \(1986\)](#) go further, applying procedure and proverb (“... *in three generations*”) to estimates from the *empirical* literature. Their striking conclusion is that differences between families and the prevalence of poverty tend to disappear within few generations.² Finally, the idea that regression implies *iterated* regression appears quite natural in a linear regression context. Indeed, such belief turns out to be a common statistical fallacy, arising frequently also in other economic literatures and disciplines.³

¹The early empirical literature focuses on occupational mobility and is mostly concerned with the (related but distinct) question if grandparents have a direct causal influence on their grandchildren; exemplary is [Warren and Hauser \(1997\)](#), who also summarises earlier studies.

²This conclusion seems conflicting with the observation that some groups, such as black Americans, experience persistent economic disadvantage. Becker and Tomes (1986) interpret this observation not as a evidence against the extrapolation procedure from intergenerational equations as such, but instead argue that those equations could differ for Blacks (e.g. p. 28).

³For example, [Bernard and Durlauf \(1996\)](#) examine tests for the *convergence hypothesis* of the neo-classical growth model, showing that a negative slope coefficient in a cross-country regression of growth rates on initial levels of output does not suggest that poorer economies tend to fully catch up to richer

In this note I present various simple models of intergenerational transmission to illustrate why iteration-based procedures are unlikely to approximate the true relation between intergenerational and multigenerational mobility. Starting from a baseline model I discuss the role of indirect transmission and market luck; the multiplicity of skills; the role of grandparents; and finally the causal effect of parental income. This discussion leads to a specific hypothesis: various properties of the intergenerational transmission process imply that long-run mobility will likely be *lower*, possibly much lower, than the standard extrapolation procedure implies.

I illustrate those arguments using data on three generations from Swedish registers, but my main objective is to provide a theoretical complement to the recent wave of recent empirical studies on the subject. [Lindahl, Palme, Sandgren Massih, and Sjögren \(2014\)](#) exploit survey data on the parents, children and grandchildren of a Swedish population, and find that multigenerational persistence in income and education is severely understated by iterated parent-child estimates. Longitudinal data as used in this and other forthcoming studies ([Dribe and Helgertz, 2013](#); [Boserup, Kopczuk, and Kreiner, 2013](#)) provide an exceptional but rare opportunity to study multigenerational persistence. Other researchers thus rely on novel methods to exploit repeated cross-sections instead: [Long and Ferrie \(2013a\)](#) link individuals in British and U.S. censuses; [Collado, Ortuño-Ortin, and Romeu \(2013\)](#) exploit socioeconomic bias in the distribution of surnames in two Spanish regions; [Clark \(2013\)](#) relies on the informative content in rare surnames; and [Olivetti, Paserman, and Salisbury \(2014\)](#) on information in first names.

In contrast, little theoretical work exists on the topic. [Zylberberg \(2013\)](#) studies the inheritance of *careers*, and shows that income persistence will decay less than geometrically if mobility is high within but low between distinct blocks of careers. [Solon \(2013\)](#) extends the classic Becker-Tomes framework to study multigenerational persistence, which leads to a variant of the simultaneous equation system that I am considering here. He pays particular attention on grandparent coefficients in multigenerational regressions, showing that its signs are ambiguous if grandparents have independent causal effects on their grandchildren. I consider various other elements of the transmission process, which together lead to a specific hypothesis on the relationship between inter- and multigenerational persistence.

ones.

4.1 The Iterated Regression Fallacy

I consider the iteration procedure in some detail, as it is common not only in the intergenerational literature. The degree to which differences in socio-economic outcomes between *parents* remain among their children is often measured by the slope coefficient in a linear regression of outcome y in offspring generation t of family i on parental outcome in generation $t - 1$,

$$y_{it} = \alpha + \beta_{-1}y_{it-1} + \varepsilon_{it}. \quad (4.1)$$

If y measures log lifetime income then β_{-1} captures the *intergenerational income elasticity*, which measures the percentage differential in expected offspring income with respect to a percentage differential in parental income; a high elasticity represents low mobility. For simplicity assume stationarity, such that β_{-1} is constant over t . The arguments apply likewise in a non-stationary environment.

How does this parent-child coefficient compare with the coefficient across three or more generations, e.g. between *grandparents* and their grandchildren? The idea that the latter equals the square of the former, so that persistence declines geometrically, may appear as a natural consequence of regression: if β_{-1} captures to what degree deviations from the mean tend to be passed from parents to children then surely $(\beta_{-1})^2$ measures their expected extent after being passed twice from parents to children? Formally, one may believe that equation (4.1) can be used to rewrite the grandparent-grandchild elasticity β_{-2} as

$$\beta_{-2} \equiv \frac{Cov(y_{it}, y_{it-2})}{Var(y_{it-2})} = \frac{Cov(\beta_{-1}y_{it-1} + \varepsilon_{it}, y_{it-2})}{Var(y_{it-2})} = (\beta_{-1})^2. \quad (4.2)$$

The error lies in the last step. While ε_{it} is by construction uncorrelated to y_{it-1} , it is not necessarily uncorrelated with grandparental outcome y_{it-2} . The interpretation of equation (4.1) itself may be the source of confusion; it has no structural interpretation, nor does it represent an AR(1) or Markov process. Instead it captures a simple statistical relationship: $\beta_{-1}y_{it-1}$ is the best linear approximation (in a MMSE sense) to the conditional expectation $E[y_{it}|y_{it-1}]$.

The belief that regression toward the mean between two observations implies iterated regression between multiple observations appears to be a classic fallacy. Francis Galton himself fell fault of it (Bulmer, 2003). As noted in the introduction, iteration-based ex-

trapolations are common in the intergenerational and other economic literatures, as well as other disciplines: Nesselroade et al. (1980) discuss their prevalence in developmental psychology under the caption "*expectation fallacy*".⁴ I use the term "*iterated regression fallacy*" here to relate to the consecutive nature of intergenerational transmission and to other classic regression fallacies.⁵

But is the extrapolation error from an iteration of regression coefficients quantitatively important? I provide a brief empirical application using Swedish population and education registers, covering a 35 percent random sample of the Swedish cohorts born between 1932 and 1967 and their biological parents and children. Educational attainment (converted into years of schooling) for offspring, their parents and their grandparents can be observed in about 150,000 cases.⁶

The first row of Table 4.1 reports coefficient estimates from child-father ($\hat{\beta}_{-1} = 0.238$) and father-grandfather ($\hat{\beta}_{-1} = 0.406$) regressions. The predicted child-grandfather coefficient based on the iteration of intergenerational measures is $0.238 \times 0.406 = 0.096$ (second row), but the estimate from an actual child-grandfather regression ($\hat{\beta}_{-2} = 0.137$, third row) is more than 40 percent higher. The extrapolation error is even larger in child-mother-grandmother coefficients (second panel). Educational differences are not very persistent in Sweden, but the iteration of intergenerational coefficients substantially understates long-run persistence.⁷

4.2 Models of Inter- and Multigenerational Transmission

The iterated regression fallacy will become more evident in the light of a theoretical model. Consider a simplified one-parent one-offspring family structure for which income

⁴The fallacy can be viewed as an incorrect application of the law of iterated expectations: extrapolation by exponentiation would be reasonable if $E[y_{it}|y_{it-2}] = E[E[y_{it}|y_{it-1}]|y_{it-2}]$, which however only holds if y_{it} follows a Markov process.

⁵Such as the belief that regression towards the mean implies convergence to the mean, see Friedman (1992); or the failure to account for regression to the mean in comparisons over time (Jerrim and Vignoles, 2012, discuss a recent example).

⁶Schooling is not observed for cohorts before 1911 and becomes increasingly right-censored in cohorts after 1972. I thus restrict my sample to individuals born 1966-1972 and their parents and grandparents.

⁷These findings are consistent with Lindahl, Palme, Sandgren Massih, and Sjögren (2014), who provide more comprehensive evidence.

and intergenerational transmission are governed by

$$y_{it} = \rho e_{it} + u_{it} \quad (4.3)$$

$$e_{it} = \lambda e_{it-1} + v_{it}, \quad (4.4)$$

such that income y_{it} depends on human capital e_{it} (according to *returns* ρ), which is partially inherited within families (according to *heritability* λ). I use the term heritability in a wide sense, representing not only genetic but also other causal pathways of transmission from parents to children (e.g. parental upbringing). The noise terms u_{it} and v_{it} represent market and endowment luck, and are assumed to be uncorrelated with each other and past values. To simplify the presentation assume throughout that variables are measured as trendless indices with mean zero and variance one, such that slope parameters can be interpreted as correlations; and further that those indices measure favourable traits that are not negatively correlated within families, such that all parameters are non-negative. The parameter ρ then measures the fraction of income that is explained by inheritable own characteristics, as opposed to factors or events outside of individual control; for example, $\rho = 1$ implies that income differences are fully explained by own characteristics. The i subscript is dropped in the subsequent analysis.

Given equations (4.3) and (4.4), and the assumption that all variances are unity, the intergenerational elasticity equals

$$\begin{aligned} \beta_{-1} &= Cov(y_t, y_{t-1}) \\ &= \rho^2 \lambda, \end{aligned} \quad (4.5)$$

and the elasticity across three generations instead equals

$$\begin{aligned} \beta_{-2} &= Cov(y_t, y_{t-2}) \\ &= \rho^2 \lambda^2. \end{aligned} \quad (4.6)$$

The extrapolation error from exponentiation of the parent-child elasticity equals

$$\begin{aligned} \Delta &= (\beta_{-1})^2 - \beta_{-2} \\ &= (\rho^2 - 1)\rho^2 \lambda^2, \end{aligned} \quad (4.7)$$

which is negative if $0 < \rho < 1$ and $0 < \lambda < 1$, that is as long as income is not perfectly determined by human capital, and human capital is not perfectly inherited within families. The extrapolation error Δ will be large when ρ is small relative to the degree of heritability captured by λ .

4.2.1 Indirect Transmission

This simple model illustrates that multigenerational coefficients cannot be recovered by the iteration of parent-child coefficients. Why do such extrapolations *overstate* mobility here? The result stems from the interplay of causal mechanisms between and within generations. Both the imperfect inheritability of traits between ($\lambda < 1$) and the imperfect determination of incomes by those traits within generations ($\rho < 1$) decrease the intergenerational persistence of income. But regression beyond two generations depends only on the heritability parameter: persistence equals $\beta_{-2} = \beta_{-1}\lambda$ across three generations, $\beta_{-3} = \beta_{-1}\lambda^2$ across four generations, and so on. The intuition is simple: traits are inherited multiple times, but they are only once transformed into income for each generation.

The underlying assumptions should be uncontroversial. We know that at least part of the intergenerational transmission of income occurs via indirect mechanisms, for example through genetic inheritance or parental upbringing. And we do not expect individuals with equivalent levels of human capital to have exactly equal incomes, perhaps because workers trade income for non-pecuniary aspects, or because factors outside of individual control drive a wedge between skill and income.⁸ The extrapolation error can be substantial even if the role of such external factors or “market luck” is modest. For example, exponentiation of an intergenerational elasticity of 0.5 implies $\{\beta_{-1}, \beta_{-2}, \beta_{-3}\} = \{0.5, 0.25, 0.125\}$ and thus rapid regression to the mean. But if $\rho = 0.8$ (market luck explains about a third of the cross-sectional variance in income) then $\lambda \approx 0.78$ and $\{\beta_{-1}, \beta_{-2}, \beta_{-3}\} = \{0.5, 0.39, 0.31\}$, implying substantial long-run persistence of economic status within families.

The gap between extrapolated and actual long-run mobility will tend to rise if the transmission of income is less direct. Assume that human capital is not directly trans-

⁸This wedge may be sizeable – earnings regressions only explain a fraction of the variation in the dependent variable, even when the list of regressors is large (e.g. [Zax and Rees, 2002](#)); monozygotic twins have substantially different earnings even while their genetic and early family background are similar; and various economic literatures show that events outside of individual control (such as occupation-, region-, or firm-specific demand shocks) affect incomes.

mitted within families either, but that parents bequeath certain traits a_t (according to heritability π), which in turn affect human capital e_t (according to transferability μ),⁹ and thus

$$y_t = \rho e_t + u_t \quad (4.8)$$

$$e_t = \mu a_t + v_t \quad (4.9)$$

$$a_t = \pi a_{t-1} + w_t. \quad (4.10)$$

The parent-child and grandparent-child elasticities are then equal to $\beta_{-1} = \rho^2 \mu^2 \pi$ and $\beta_{-2} = \rho^2 \mu^2 \pi^2$. Consider parameterizations that yield the same parent-child elasticity as the two-layers model, which requires $\lambda = \mu^2 \pi$. The extrapolation error then equals $\Delta = (\rho^2 - \frac{1}{\mu^2}) \rho^2 \lambda^2$. Comparison to equation (4.7) reveals that this error will be larger in the three- than in the two-layers model iff $\mu < 1$.

Various implications follow. First, long-run income mobility will be smaller the more intergenerational mobility is attributable to market luck instead of low heritability of traits. Second, policies and institutions may mask inequality only temporarily. For example, a track-school system that separates children by ability may increase the degree to which differences in child ability lead to differences in human capital (an increase of μ in the three-layers model) and thus decrease intergenerational mobility. But long-run mobility may be less affected if the heritability of those abilities remains unchanged. Third, the degree to which cross-country differences in parent-child mobility extend to the long run depends on if those differences are due to variation in the heritability or the the transferability of endowments.¹⁰ Clark (2013) pushes this idea further, arguing that long-run mobility is instead closer to a universal constant across countries and time.

4.2.2 Identification from Multigenerational Data

I focus here on the implications of different causal processes for the extent of multigenerational mobility. But that relationship is interesting also the opposite way, as it may help to identify features of the underlying causal process that are otherwise difficult to

⁹As my main intention is to capture the idea that income transmission may occur rather indirectly I abstain from specific interpretations for each layer (e.g. the lowest layer may be thought to represent genetic transmission, as in Conlisk, 1974a).

¹⁰For example, Nordic countries will be characterised by exceptional long-run mobility if their high levels of intergenerational mobility are caused by policies and institutions that decrease the heritability of traits, less so if they are due to policies that interfere with the formation of market prices for those traits.

capture. For example, the heritability of ability might not be directly estimable, as at best we can hope to observe a noisy proxy for certain types of traits. Multigenerational data offers potentially an indirect route to identification. In the three-layers model, the slope coefficient from a regression of offspring on parent human capital equals $\beta_{-1} = \mu^2\pi$ while the slope coefficient from a child-grandparent regression equals $\beta_{-2} = \mu^2\pi^2$. The ratio β_{-2}/β_{-1} identifies thus π , while $(\beta_{-1}^2/\beta_{-2})^{1/2}$ identifies μ , from data on educational attainment alone.

Empirical implementation of distributional models is not straightforward if the environment cannot be assumed to be in steady state (see [Atkinson and Jenkins, 1984](#)). Indeed, the slope coefficient in a parent-child regression of years of schooling changes considerably over time in Sweden (see Table 4.1), predominantly due to variation in the cross-sectional variance. To abstract at least from this variation consider instead the correlation coefficient r , which is considerably more stable. Dividing the child-grandfather correlation ($\hat{r}_{-2} = 0.156$) by the average of the child-father ($\hat{r}_{-1} = 0.323$) and father-grandfather ($\hat{r}_{-1} = 0.340$) correlations yields $\hat{\pi} = 0.471$, and thus $\hat{\mu} = 0.839$.

These estimates can in turn be used to extrapolate beyond three generations. Simple iteration of the intergenerational correlation yields

$$\{r_{-1}, r_{-2}, r_{-3}, r_{-4}\} = \{0.332, 0.110, 0.037, 0.012\},$$

but the model-based procedure implies

$$\{r_{-1}, r_{-2}, r_{-3}, r_{-4}\} = \{0.332, 0.156, 0.074, 0.035\}.$$

This prediction is still flawed if the true causal process is not well captured by eqs. (4.8) to (4.10). But in contrast to the iteration of bivariate coefficients it has a conceptual justification and is consistent with data over three instead of two generations. Its validity can be tested if more than three generations are observed.

4.2.3 An Additional Factor

A second fundamental reason why multigenerational persistence of economic status may decay less rapidly in the long than in the short run relates to the multiplicity of the

transmission process. Introduce a second factor into our starting model,

$$y_t = \rho_1 e_{1t} + \rho_2 e_{2t} + u_t \quad (4.11)$$

$$e_{1t} = \lambda_1 e_{1,t-1} + v_{1t} \quad (4.12)$$

$$e_{2t} = \lambda_2 e_{2,t-1} + v_{2t}, \quad (4.13)$$

assuming that two traits are inherited from parents according to heritability parameters λ_1 and λ_2 . For simplicity also assume that the endowment luck terms v_{1t} and v_{2t} are uncorrelated, such that $Cov(e_{1t}, e_{2t}) = 0 \forall t$. Assume further that both traits affect incomes, such that $0 < \rho_1 < 1$ and $0 < \rho_2 < 1$. The parent-child elasticity then equals

$$\beta_{-1} = \rho_1^2 \lambda_1 + \rho_2^2 \lambda_2, \quad (4.14)$$

and the grandparent-grandchild elasticity equals

$$\beta_{-2} = \rho_1^2 \lambda_1^2 + \rho_2^2 \lambda_2^2. \quad (4.15)$$

The extrapolation error equals

$$\Delta = (\rho_1^2 - 1)\rho_1^2 \lambda_1^2 + (\rho_2^2 - 1)\rho_2^2 \lambda_2^2 + 2\rho_1^2 \rho_2^2 \lambda_1 \lambda_2. \quad (4.16)$$

Assume for a moment that incomes are indeed *perfectly* determined by individual traits, such that $\rho_1^2 + \rho_2^2 = 1$ and $Var(u_t) = 0$. Equation (4.16) can then be written as

$$\Delta = \rho_1^2 (\rho_1^2 - 1) (\lambda_1 - \lambda_2)^2. \quad (4.17)$$

This expression is negative for $\lambda_1 \neq \lambda_2$. In contrast to the previous models, exponentiated parent-child elasticities understate multigenerational persistence even when human capital determines incomes perfectly, as long as those traits that constitute human capital are not all equally strong inherited within families.¹¹ This result can be understood as the application of Jensen's inequality: the square of the average heritability across traits is smaller than the average of the square of those heritabilities. Inequality between families declines therefore more slowly if intergenerational income persistence stems from multiple

¹¹For example, [Anger \(2011\)](#) and [Grönqvist, Öckert, and Vlachos \(2010\)](#) study if inheritance is stronger in cognitive than in non-cognitive abilities.

causal pathways.

Highly inheritable traits explain an increasing share of the long-run persistence in income. In particular, multigenerational elasticities will never converge to zero if any characteristic is perfectly transmitted. For example, physical traits such as skin colour may be highly persistent in multi-ethnic societies if interracial marriage is rare, and may lead to persistent disadvantage of families if groups are discriminated on the labour market. The observation of high intergenerational mobility can therefore be consistent with substantial long-run persistence of economic status, provided that the multiplicity of traits is taken into account.

For the analysis of long-run mobility it is thus essential to look beyond scalar models, even if those models have proved to be useful for other questions in the literature. This applies in particular to the framework presented in [Becker and Tomes \(1979\)](#), which underlies much of the theoretical work in the literature. It contains only a scalar measure of human capital and does not capture implications from the existence of multiple transmission mechanisms. Moreover, human capital is often assumed to determine incomes perfectly in extended versions of this model, such as [Solon \(2013\)](#). I find that both the imperfect relation between skills and income and the multiplicity of those skills have important implications for long-run persistence.¹²

4.2.4 An Additional Generation

A question that has received much attention in multigenerational studies is if grandparents have a direct causal influence on their grandchildren (e.g. [Warren and Hauser, 1997](#); [Mare, 2011](#); [Long and Ferrie, 2013a](#)). Such higher-order effects are often presented as a potential explanation why multigenerational persistence might decline more slowly than at a geometric rate. But the previous sections illustrated that other properties of the transmission process lead to the same implication. From the observation that $(\beta_{-1})^2 < \beta_{-2}$ we can therefore not conclude that intergenerational transmission has a memory of more than one generation.

The intuition that such higher-order effects raise long-run persistence is of course

¹²These findings support arguments made by [Goldberger \(1989\)](#), who notes that an explicit consideration of utility maximisation behaviour of parents (as in [Becker and Tomes, 1979](#)) to motivate “mechanical” transmission equations may provide little additional implications but distract from the assumed properties of those equations. The Becker and Tomes model leads to transmission equations that are simplified compared to earlier models in the literature, which did contain noise terms to capture market luck and multiple inheritance mechanisms (e.g. [Conlisk, 1969](#)).

correct. To see this assume that offspring human capital depends on both parents and grandparents, such that equation (4.4) becomes

$$e_t = \lambda_{-1}e_{t-1} + \lambda_{-2}e_{t-2} + v_t, \quad (4.18)$$

with $\lambda_{-2} > 0$. Assuming stationarity the parent-child elasticity equals

$$\beta_{-1} = \rho^2 \left(\frac{\lambda_{-1}}{1 - \lambda_{-2}} \right), \quad (4.19)$$

Consider parameterizations that yield the same intergenerational elasticity as the previous model, such that $\lambda = \lambda_{-1}/(1 - \lambda_{-2})$. The grandparent-grandchild elasticity,

$$\beta_{-2} = \rho^2 \lambda^2 + \rho^2 \lambda_{-2}(1 - \lambda^2), \quad (4.20)$$

is then greater than the respective elasticity in the baseline model (assuming $\rho > 0$ and $\lambda < 1$). This simple example does not illustrate the various ways how grandparents may influence their grandchildren, but it illustrates that such influence strengthens multigenerational relative to intergenerational persistence.¹³

How can we distinguish such higher-order from other causal mechanisms? One potential strategy is to find quasi-exogenous variation in the exposure to certain family members in a careful research design (Adermon, 2013). But simpler methods may be sufficient to bound their magnitude. Note that given eq. (4.18), the slope coefficient in linear regression of child on grandparent human capital equals

$$\beta_{-2,e} = \text{Cov}(e_t, e_{t-2}) = \lambda_{-2} + \lambda_{-1}\text{Cov}(e_{t-1}, e_{t-2}). \quad (4.21)$$

The coefficient in a regression of child on grandparent characteristics may have a positive coefficient either because grandparents have a direct effect on grandchildren ($\lambda_{-2} \neq 0$) – or because grandparent and parent characteristics are correlated, and children are affected by the latter. These two channels are in my simple model identified by the coefficients in a regression of child on parent *and* grandparent human capital. But in practice we cannot be sure that all relevant parent characteristics are included; we do not know if a positive coefficient on grandparents reflects direct causal effects or an omitted variable

¹³Mare (2011) and Solon (2013) suggest various channels via which grandparents may affect their grandchildren.

bias.

However, we can illustrate the potential magnitude of this bias by conditioning on exceedingly many characteristics of the parent generation. Column (1) of Table 4.2 reports estimates from a regression of offspring on fathers' and (paternal) grandfathers' years of schooling. The estimated coefficient on grandfathers' schooling is sizeable and statistically significant. However, already the inclusion of mothers' years of schooling reduces this estimate by nearly one half (column 2); a large fraction of the grandparent coefficient reflects that fathers with highly-educated grandfathers tend to have highly-educated partners, and the latter have a more direct relation with child outcomes. Controlling for parental income (column 3), allowing for a more flexible functional form by including schooling levels as indicator variables (column 4), or including schooling for all grandparents (column 5) reduces the grandfather coefficient to a precisely estimated zero.

For a subset of the fathers we observe detailed information on cognitive and non-cognitive ability from military enlistment tests. This subsample is small and quite peculiar, as test scores are observed only for the youngest parents in my sample. It is nevertheless interesting that their inclusion pushes the grandfather coefficient below zero (column 6).

These results do not suggest that grandparents have no direct effects on their grandchildren.¹⁴ They however suggest that a consideration of parent-child transmission processes may often be sufficient even if our objective is to understand multigenerational persistence.¹⁵ The observed deceleration of regression to the mean beyond two generations is more plausibly explained by the multiplicity and indirectness of parent-child transmission processes.

Parental Investment

All previous results point to “excess persistence”, to the conclusion that extrapolated intergenerational elasticities understate long-run persistence. But we can certainly think of circumstances in which the opposite holds, for which I will give one example.

¹⁴First, we are considering only distributional, not mean effects. Second, one might expect a *negative* grandparent coefficient in the absence of direct grandparental effects, as explained in Solon (2013). Finally, grandparents may play a more important role in other populations or in other outcomes, such as wealth.

¹⁵Warren and Hauser (1997), based on a smaller sample but more comprehensive analysis, come to a similar conclusion.

Assume that parental income or economic status have a causal effect on offspring; for example indirectly through parental investments in offspring human capital, or more directly through reputation or networking effects on the labour market. Consider the first case, such that equations (4.3) and (4.4) change into

$$y_t = \rho e_t + u_t \quad (4.22)$$

$$e_t = \theta y_{t-1} + \eta e_{t-1} + v_t. \quad (4.23)$$

The parent-child and grandparent-grandchild elasticities then equal

$$\begin{aligned} \beta_{-1} &= \rho\theta + \rho^2\eta \\ \beta_{-2} &= (\rho\eta + \rho^2\theta)(\rho\eta + \theta). \end{aligned}$$

Consider again parameterizations that yield the same level of β_{-1} , which requires $\eta < \lambda$ (assuming $\rho > 0$ and $\theta > 0$). The extrapolation error,

$$\Delta = (\rho^2 - 1)\eta\beta_{-1}, \quad (4.24)$$

is smaller than the error in our first model (which equals $(\rho^2 - 1)\lambda\beta_{-1}$), but it will still be negative. Our previous findings still hold when parental income affects offspring human capital.

Now instead assume that parental income has a direct effect on offspring income that is independent of offspring characteristics, such that equations (4.3) and (4.4) change into

$$y_t = \phi y_{t-1} + \tau e_t + u_t \quad (4.25)$$

$$e_t = \lambda e_{t-1} + v_t. \quad (4.26)$$

The parent-child and grandparent-grandchild elasticities then equal

$$\begin{aligned} \beta_{-1} &= \phi + \frac{\tau^2\lambda}{1 - \phi\lambda} \\ \beta_{-2} &= \phi^2 + \frac{\tau^2\lambda}{1 - \phi\lambda}(\phi + \lambda), \end{aligned}$$

The extrapolation error equals

$$\Delta = \left(\frac{\tau^2 \lambda}{1 - \phi \lambda} \right)^2 + (\phi - \lambda) \frac{\tau^2 \lambda}{1 - \phi \lambda}. \quad (4.27)$$

which in contrast to the previous examples may be positive, in particular if $\phi > \lambda$. In this model, income is affected by both parental income and ability, but offspring ability is affected exclusively by parental ability. While parent-child persistence may be strongly affected by the direct effect of income ϕ , long-run persistence will be dominated by the heritability of ability λ . If the former is larger than the latter we have a system in which multigenerational persistence is weaker than exponentiation of β_{-1} implies.

We may expect that the causal effect of parental income is small (Björklund and Jäntti, 2009), and that at least part of it is indirect, for example through parental investments in child human capital. The case $\phi > \lambda$ seems thus of less practical relevance. Still, the model has interesting implications for cross-country differences in short- and long-run mobility.¹⁶ The direct effect of parental income captured by ϕ will tend to be larger if credit constraints are more important. Eq. (4.27) then implies that extrapolation from parent-child correlations understates long-run persistence more in those countries in which credit constraints play less of a role.

These models illustrate that different beliefs about causal pathways of transmission are consistent with different expectations about long-run mobility. The belief that children from affluent families tend to fare better mainly because inherited traits and parental investment raise their productive abilities is consistent with the expectation that long-run mobility is lower than the iteration of intergenerational coefficients implies. But some authors emphasise the importance of genetic inheritance and parental investment while challenging the significance of low intergenerational mobility estimates on the grounds that they nevertheless imply high long-run mobility (Mankiw, 2006). The opposite argument applies if one believes that income persistence stems mainly from mechanisms that are unrelated to individual productivity. If income persistence is *only* due to the direct influence of parental income then its decline over generations is indeed geometrically, and even low levels of intergenerational mobility would imply rapid multigenerational regression to the mean.

¹⁶I thank an anonymous referee for the following observations.

4.3 Conclusions

For lack of direct evidence, predictions on the degree of long-run mobility across multiple generations are routinely derived by extrapolation from intergenerational evidence. But an iteration of parent-child correlations implies high long-run mobility even when intergenerational mobility is low – suggesting that the extensive literature measuring such correlations is of little consequences for the distribution of socio-economic characteristics beyond two generations.

In this note I studied various elements of the intergenerational transmission process that lead to a different conclusion: the persistence of socio-economic differences in status is likely to be higher, perhaps much higher than the iteration of parent-child correlations implies.

The idea to iterate slope coefficients appears quite natural in a linear regression context, and this *iterated regression fallacy* is widespread not only in the intergenerational but also in other economic literatures and disciplines. I first illustrated why iteration-based extrapolation procedures are unlikely to provide a good approximation of multigenerational persistence. I provided empirical support for this argument using data on educational attainment across three generations from Swedish registries. Regression to the mean is comparatively strong in Sweden, but its rate slows indeed substantially after two generations.

I considered then how various elements of the transmission process affect the relation between inter- and multigenerational mobility. I discussed the role of direct and indirect pathways of transmission; of the multiplicity of skills; of higher-order causal effects; and the role of parental income. Regression to the mean slows over generations if factors that are orthogonal to individual characteristics explain some fraction of the variation in socio-economic outcomes. The multiplicity of skills also matters, as highly inheritable traits explain an increasing share of socio-economic persistence across generations. Moreover, multiplicity provides a simple explanation why groups can suffer from persistent economic disadvantage even when parent-child mobility is high.

The recent literature has been particularly concerned with the question if grandparents have a direct causal influence on their grandchildren. I argued that other properties of the transmission process are more important for understanding long-run persistence in socio-economic outcomes. I used the Swedish data for a simple illustration: while

all coefficients in a regression of child on parent and grandparent schooling tend to be positive, the coefficient on grandparent schooling vanishes quickly and may even turn negative when we include a wider set of parental controls.

Questions on mobility across multiple generations are closely related to questions on the causal pathways of transmission. This note focused on the implications that different causal channels have for the extent of multigenerational mobility, but the relation is interesting both ways. I illustrated how a comparison of intergenerational and three-generation coefficients may help to identify features of the underlying causal process and lead to better extrapolations from the available intergenerational evidence.

Tables and Figures

Table 4.1: Inter- and Multigenerational Persistence in Educational Attainment

	Child	Father	Grandfather
two generations	0.238*** (0.002)	0.406*** (0.004)	
three gen. (prediction)		0.096*** (0.001)	
three gen. (actual)		0.137*** (0.003)	
	Child	Mother	Grandmother
two generations	0.267*** (0.002)	0.301*** (0.005)	
three gen. (prediction)		0.080*** (0.002)	
three gen. (actual)		0.152*** (0.004)	

Notes: Slope coefficients from separate regressions of years of schooling of offspring on years of schooling of family member in older generation. N=145,590 observations for panel A (fathers/grandfathers), N=156,847 for panel B (mothers/grandmothers). Standard errors (in parentheses) are clustered by family.

Table 4.2: The Grandparent Coefficient in Educational Persistence

	Years of schooling - Child					
	(1)	(2)	(3)	(4)	(5)	(6)
Parents:						
schooling father	0.222*** (0.0025)	0.159*** (0.0026)	0.135*** (0.0027)	saturated	saturated	saturated
schooling mother		0.182*** (0.00289)	0.169*** (0.0029)	saturated	saturated	saturated
income father			0.546*** (0.0171)	0.461*** (0.0169)	0.407*** (0.0245)	0.191*** (0.068)
income mother			-0.0176 (0.0095)	-0.0021 (0.0094)	0.0298* (0.0152)	0.120** (0.055)
ability father	-	-	-	-	-	x
Grandparents:						
schooling grandfather (paternal)	0.0456*** (0.0031)	0.0259*** (0.0030)	0.0183*** (0.0030)	0.0083** (0.0030)	0.0029 (0.0047)	-0.0132 (0.0164)
schooling grandmother (paternal)					0.0069 (0.0060)	
schooling grandfather (maternal)					0.0061 (0.0046)	
schooling grandmother (maternal)					0.0069 (0.0059)	
# obs.	104,904	104,904	104,904	104,904	47,797	2,789

Notes: Slope coefficients from separate regressions of years of schooling of offspring on characteristics of parents and grandparents. Standard errors (in parentheses) are clustered by family.

Chapter 5

The Impact of Immigration on Local Labor Markets:

Evidence from the Opening of the German-Czech Border

5.1 Introduction

The effect of immigration on wages and employment of native workers has been studied in numerous papers, yet there is little consensus about whether immigration has adverse impacts, and what is the most suitable methodology that should be adopted for analysis. In this paper we revisit the question of how an immigration induced labor supply shock affects wages and employment of native workers.

We argue that analysis of wage effects alone or without the consideration of group-specific labor supply responses – the type of analysis performed by most works in this area – may lead to misleading conclusions on how immigration affects different types of workers. This is because the labor supply responses of different groups of resident workers to an immigration induced labor supply shock may differ across subgroups of workers.. One key innovation of this paper is therefore to allow for different labor supply elasticities of different groups of resident workers, and to estimate these in an analysis that investigates wage and labor supply jointly.

A second contribution of our work is that we distinguish between different *types* of labor supply responses. The overall response of any group of resident workers to a labor supply shock will consist of individuals leaving, or not entering employment in the affected area. These responses may in turn be due to movements into or out of employment within, or geographic movements between local labor markets. We provide evidence on the magnitude of each type of response, and show how the relative importance of each adjustment channel varies across groups.

Our analysis is based on a unique experiment that took place in Germany fourteen months after the fall of the Berlin wall, when the German government introduced a commuting scheme that allowed workers from the neighboring Czech Republic to seek employment in German municipalities along the German-Czech border. While these workers were allowed to work in Germany, they were not granted residence, which forced them to commute between their home country and their workplace in Germany. This resulted in an almost ideal labor supply shock that affected only eligible municipalities, and which was exogenous to local conditions on the labor market. The supply shock was unexpected, sudden and of considerable magnitude: within two years the average share of Czech workers increased from close to zero

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

to more than three percent across all eligible municipalities, and to about 10 percent in areas close to the border. While the nature of this experiment is not dissimilar to David Card's Miami boat lift (Card, 1990), the requirement to commute on a daily basis provides us with additional exogenous variation for the intensity with which different municipalities were affected by this policy, in form of measures of distance from the border. Moreover, the policy generated a pure labor supply shock that was less contaminated by possible demand effects, as commuting workers did not live and consume in the affected areas. Our experiment thus allows for a more direct test of the predictions of factor demand models that are at the core of theoretical work in the literature (Borjas, 2009).

Our empirical investigation is based on a unique administrative data set that covers the totality of workers subject to social security contributions in Germany. This data reports not only on a wide range of socio-economic characteristics of workers, but contains also citizenship information, which allows us to identify Czech workers who entered Germany in response to the policy. Moreover, as we have longitudinal information on each worker, we are able to precisely investigate the various channels of employment responses of individual workers. The continuous updates on employment status, and at least yearly updates on wages, also allow us to study in far more detail the timing of wage and employment responses. Thus, the combination of a unique policy, a clean identification strategy, and unique individual and longitudinal data on each worker who was potentially affected allows us to extend the existing literature into various new directions, and to reconcile some of the apparently contradictory findings that have puzzled researchers in the past. Moreover, the longitudinal nature of our data and the sudden onset of the experiment allow us to perform very targeted falsification tests and thus to carefully probe the robustness of our identification strategy.

We find that the large inflow of Czech workers over a two years period had a negative effect on overall wages and led to large employment responses among native workers. Both responses were remarkably rapid, with the wage response preceding the full employment response. As expected, the inflow of predominantly unskilled Czech workers led to larger employment and wage losses among unskilled native workers. We further demonstrate that an analysis based on repeated cross-sectional data, even if comparing wages just before and shortly (i.e. one year) after the labor supply shock, fails to pick up much of the wage effects,

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

due to low-productivity workers being those that leave the affected area (or do not enter it in the first place) in response to the shock.

We further show that labor supply responses differ strongly between different skill groups and between workers of different age, implying heterogeneity in the response to wage changes across skill and demographic groups that are explained by different labor supply elasticities. We demonstrate that approaches that focus on responses in *relative* wages, induced by relative labor supply shocks between skill groups, may lead to misleading conclusions when heterogeneity in resident workers' labor supply responses is ignored.

We then investigate in detail the relative importance of various types of employment adjustment. We show that the employment responses of native workers in municipalities affected by an inflow of Czech workers decreases predominantly through reductions in *inflows* into local employment, while *outflows* from the existing native workforce are much smaller in magnitude. We also find that inflows adjust faster than outflows, which may explain why employment levels adjust so rapidly. This provides an entirely new interpretation of labor supply shock-induced employment responses, not only for such shocks being induced through migration, but more generally.

We further investigate the margins of adjustment along another dimension, distinguishing between geographic movements of native workers between areas and employment and non-employment. Native employment adjusts mostly through reduced inflows from and increased outflows into non-employment. Interesting is further that we find the importance of those channels to vary across subgroups of workers, with for instance older workers reacting strongly through outflows into non-employment, while the response of younger workers is predominantly driven by reductions in inflows from both non-employment and other areas.

To structure our empirical analysis and to better understand our estimates, we commence with a simple model of production as is standard in the migration literature. We extend that model, by allowing different skill groups to exhibit different labor supply elasticities, and we allow for additional heterogeneity between demographic groups within the same skill group. Based on that model, we show that heterogeneity in responses may lead to “perverse” wage effects, with relative wages of that skill group increasing that experiences the larger labor supply shock, if only the labor supply elasticity of this group is large enough.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

Moreover, if different groups of workers who compete in the same labor market segment respond differently to a wage changes, then any skill group specific labor supply elasticity will be determined by the labor supply response of those workers who respond most elastically, muting wage effects of all other workers in that skill group

Our work relates to different literatures. We contribute to the literature on wage impacts of immigration (Borjas 1999, Card 2009, Aydemir and Borjas 2011, Dustmann et al. 2013) by showing that a pure wage analysis may not capture the full impact of immigration even if wages are measured shortly after the immigration-induced supply shock, as employment responses of natives are too rapid. We further show how approaches that rely on the classification of immigrants into skill and demographic cells (Borjas et al. 1997, Borjas 2003, Borjas et al. 2010, Manacorda et al. 2012, Ottaviano and Peri 2012) are susceptible to response heterogeneity among natives across those cells.

Our paper is also related to the literature that discusses possible geographic outflows of natives as a response to immigration induced labor supply shocks (see e.g. Borjas et al. 1997, Card 2001, Card and DiNardo 2000), which may dissipate any wage effects across the economy. While that literature assesses in particular if native workers *leave* areas in response to migration, we consider here that individuals may instead not enter employment in that market. We find that compared to within-labor market adjustments, movements across individual labor markets are small, even for the fine geographical definitions we use in this study. More importantly, we also find that both responses, rather than through outflows, are through reducing potential inflows into employment in an affected area.

The idea that workers may differ in their labor supply elasticities suggests that different groups of workers may be differently affected by labor supply shocks, where those workers who react most elastically contribute more strongly to wage effects being muted, thus “shielding” other resident workers from the impact of immigration-induced supply shocks. This idea is in spirit similar to findings in a recent paper by Cadena and Kovak (2013), who find that unskilled immigrants react most strongly to labor *demand* shocks during the great recession, leading to native workers being insulated from such shocks in cities with large Mexican born populations. It also relates to findings in a recent paper by Smith (2012), who shows that unskilled immigration leads to different employment outcomes across different demographic groups, which he interprets as being partly due to differences in responsiveness

across groups to wage changes. We contribute to this literature by providing a theoretical structure, and investigating such response heterogeneity in a setting where we are able to precisely identify wage and employment effects, and the various channels via which employment adjust, across groups of workers.

5.2 An Equilibrium Model with Heterogeneous Labor Supply

We first set out a simple model which links immigration-induced labor supply shifts to employment and wage responses of natives in that labor market, and develops the key implications for the empirical analysis. One distinguishing feature of our model compared to existing models is that we allow local labor supply elasticities to vary across skill groups.

2.1 Baseline Model

Production: Suppose that output Q in a specific area is produced by combining labor L and capital K according to a Cobb-Douglas production function:

$$Q = AK^\alpha L^{1-\alpha}. \quad (1)$$

Labor L is a CES aggregate of unskilled (U) and skilled (S) labor L_g , $g = U, S$:

$$L = \left[\theta_U L_U^\beta + \theta_S L_S^\beta \right]^{\frac{1}{\beta}}, \quad (2)$$

where $\theta_U + \theta_S = 1$, and the elasticity of substitution between the two skill groups equals $\sigma = \frac{1}{1-\beta}$, with $\beta \leq 1$.

Labor supply: In each skill group g , total labor supply in the area is the sum of labor supply of natives (or incumbents), L_g^N , and immigrants (or entrants), L_g^I . Suppose that immigrants and natives are perfect substitutes in production and that there are no immigrants in the base period (which corresponds to our empirical setting), so that $L_g = L_g^N$. Denote by N_g the (fixed) number of natives who could potentially supply labor to the local labor market. The local labor supply function for skill group g is then

$$L_g = L_g^I + L_g^N = L_g^I + N_g f_g(w_g, \mathbf{w}'_g), \quad (3)$$

where we assume that immigrants (i.e., new entrants) supply their labor inelastically, whereas the local labor supply of natives (i.e., incumbents) depends on skill-specific wages in the market under consideration (w_g) and other local labor markets (\mathbf{w}'_g).

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

The local labor market elasticity for natives, which we allow to vary across skill groups, is given by $\eta_g = \frac{\partial N_g f_g(\cdot)}{\partial w_g} \frac{w_g}{N_g f_g(\cdot)}$. It summarizes various potential adjustment mechanisms, such as the internal migration of workers between areas, or entries into and exits from the labor force. We comment on both these dimensions, as well as why the elasticity may vary across groups, below and in Section 5.

Assume that (as it is the case in our empirical application) the local labor market under consideration is small relative to the national labor market. In consequence, the change in equilibrium wages (and native employment) in other areas is negligible. Hence, totally differentiating the expression in (3), and dividing both sides by equilibrium employment in the base period (recall that in the base period, $L_g = L_g^N$) results in

$$d \log L_g = \frac{dL_g^I}{L_g^N} + \eta_g d \log w_g \quad (4a)$$

This expression decomposes the percentage increase in the labor supplied to the local labor market into an exogenous increase caused by the inflow of immigrants, $\frac{dL_g^I}{L_g^N}$, and an endogenous response of natives, $\eta_g d \log w_g$ – which is stronger the larger their local elasticity of labor supply. Let $\pi_g^N = \frac{L_g^N}{L_U^N + L_S^N}$ and $\pi_g^I = \frac{L_g^I}{L_U^I + L_S^I}$ denote the employment shares of natives and immigrants of skill group g (in head counts), to rewrite the expression above as

$$d \log L_g = \frac{\pi_g^I}{\pi_g^N} dI + \eta_g d \log w_g \quad (4b)$$

where $dI = \frac{d(L_U^I + L_S^I)}{(L_U^N + L_S^N)} = \frac{dL^I}{L^N}$ is the total labor supply shock to the local labor market induced by migration relative to native equilibrium employment in the base period (in head counts). Differentiating the expression in (2) and substituting expression (4b) for $d \log L_g$ yields¹

$$d \log L = \Pi dI + s_U \eta_U d \log w_U + s_S \eta_S d \log w_S \quad (5)$$

where $\Pi = s_U \frac{\pi_U^I}{\pi_U^N} + s_S \frac{\pi_S^I}{\pi_S^N}$ is the weighted average of the relative density of immigrants across skill groups, and where weights s_U and s_S are the contribution of labor type g to the total labor aggregate.

¹ It follows from (2) that $d \log L = \frac{\theta_U L_U^\beta}{[\theta_U L_U^\beta + \theta_S L_S^\beta]} d \log L_U + \frac{\theta_S L_S^\beta}{[\theta_U L_U^\beta + \theta_S L_S^\beta]} d \log L_S$, so that $s_g =$

$$\frac{\theta_g L_g^\beta}{[\theta_U L_U^\beta + \theta_S L_S^\beta]}.$$

Labor demand: Assuming that markets are competitive and normalizing the price of the output good to 1, the first order conditions for the cost-minimizing input choice give the equilibrium input prices as:

$$\log w_g = \log[(1 - \alpha)A] + \alpha[\log K - \log L] + \log \theta_g + (\beta - 1)[\log L_g - \log L] \quad (6a)$$

$$\log r = \log \alpha A + (\alpha - 1)[\log K - \log L] \quad (6b)$$

Assume that the supply of capital is given by

$$\log K = \frac{1}{\lambda} \log r; \lambda \geq 0 \quad (7)$$

where $\lambda \geq 0$ is the inverse of the elasticity of capital with respect to its price r . Equate equations (5b) and (6) and solve for $\log K$ to obtain $\log K = \frac{\log \alpha A}{1 - \alpha + \lambda} - \frac{\alpha - 1}{1 - \alpha + \lambda} \log L$. Substituting for $\log K$ in (5a) using the expression above yields

$$\log w_g = \Gamma + (\beta - 1) \log \left(\frac{L_g}{L} \right) + \phi \log L \quad (8)$$

where $\Gamma = \log[(1 - \alpha)A] + \alpha \frac{\log \alpha A}{1 - \alpha + \lambda} + \log \theta_g$ and where $\phi = -\frac{\alpha \lambda}{1 - \alpha + \lambda}$ is the slope of the aggregate labor demand curve. Totally differentiating the expression in (8) results in

$$d \log w_g = \phi d \log L + (\beta - 1)(d \log L_g - d \log L) \quad (9)$$

In equilibrium, supply must equal demand, and the supply curves given by equation (4b) and the demand curves given by equation (9) determine the skill-specific and aggregate wages and employment in the local labor market.

The Impact of labor supply shocks on wages and employment: Plugging in the expressions in (4b) and (5) for $d \log L_g$ and $d \log L$ into equation (9) and solving for $d \log w_g$, we obtain the change in equilibrium wages of skill group g as a response to a labor supply shock induced by immigration (see Appendix A for a derivation):

$$\begin{aligned} d \log w_g &= \frac{(\beta - 1) \left[\frac{\pi_g^I}{\pi_g^N} (1 - \phi \eta_{g'}) - \Pi \left(1 - \frac{\phi}{\beta - 1} \right) \right]}{1 - (\beta - 1) [\eta_g (1 + s_g \phi) + \eta_{g'} (1 + s_{g'} \phi) - \eta_g \eta_{g'} (1 + \phi) (\beta - 1)]} dI \end{aligned} \quad (10)$$

where $\eta_{g'}$ is the local labor market elasticity of the *other* skill group, and $\phi = \frac{\phi}{\beta - 1} - 1$. Local employment responses of natives in the area follow straightforwardly from the labor supply function:

$$dlogL_g^N = \eta_g dlogw_g \quad (11)$$

As $\beta \leq 1$, the denominator in (10) will always be positive. The numerator is the difference in the relative density of immigrants to natives in skill group g , $\frac{\pi_g^I}{\pi_g^N}$, and the (weighted) average of these densities in the different skill groups, Π , both weighted by expressions that depend on the elasticity of capital supply (ϕ), and the supply elasticity of the other labor type ($\eta_{g'}$). Thus, the impact of a supply shock on native wages will be negative for skill group g if the weighted intensity of immigration in that skill group (first term in brackets) exceeds an appropriately weighted average of immigration intensity across all skill groups (second term in brackets).

In the standard case of a homogenous local labor supply elasticity (i.e., $\eta_U = \eta_S = \eta$), equations (10) and (11) imply that wages and employment of the skill group which experiences the larger migration-induced supply shock (i.e. the group for which $\pi_g^I/\pi_g^N > \Pi$) decline relative to wages and employment of the other group. Wage effects will be the smaller, and employment effects the larger, the larger the labor supply elasticity. Moreover, wages and local employment of both skill groups may decline if capital is not fully elastic ($\phi \neq 0$).²

If, in contrast, the local labor supply elasticity varies across groups, then it is possible that wages of the skill group for which immigration is relatively intensive *increase* relative to the other skill group. To see this, consider the *relative* wage effects:

$$dlogw_S - dlogw_U = \frac{(\beta - 1) \left[\frac{\pi_S^I}{\pi_S^N} (1 - \phi \eta_U) - \frac{\pi_U^I}{\pi_U^N} (1 - \phi \eta_S) \right]}{1 - (\beta - 1)[\eta_S(1 + s_S\phi) + \eta_U(1 + s_U\phi) - \eta_U\eta_S(1 + \phi)(\beta - 1)]} dI$$

Suppose that migration is predominantly unskilled (i.e., $\frac{\pi_S^I}{\pi_S^N} < \frac{\pi_U^I}{\pi_U^N}$) and that local labor supply of the unskilled is elastic relative to labor supply of the skilled (i.e., η_U is large relative to η_S). In this case, the relative employment effect is amplified, whereas the relative wage effect is muted, compared to the case of a homogenous local labor supply elasticity. Provided that

² In the case of homogenous labor supply elasticities, expression (8) simplifies to

$$dlogw_g = \frac{\phi\pi}{1 - \phi\eta} dI + \frac{(\beta - 1)}{1 - \eta(\beta - 1)} \left(\frac{\pi_g^M}{\pi_g^N} - \pi \right) dI.$$

This expression does not only highlight that wages of the skill-group for which $\pi_g^I/\pi_g^N > \pi$ decline relative to the other skill group, but also shows that wages of both the unskilled and skilled may decline if capital is not fully elastic – see the first term in the expression above, which pulls down wages for both skill groups by the same amount, the more so the less elastic capital supply. If $\eta = 0$, the expression reduces to the case discussed in Dustmann et al. (2013).

capital is not fully elastic ($\varphi < 0$) and some skilled migrants enter the local labor market ($\pi_S^I > 0$), it is possible that wages of unskilled increase relative to wages of skilled.

If on the other hand the local labor supply elasticity of the unskilled is small relative to that of the skilled, the difference in wage adjustments between the unskilled and skilled will be amplified, and the differences in employment adjustments muted, compared to the case of a homogenous local labor supply elasticity. It is now possible that employment of the unskilled increases relative to the skilled, even though more unskilled than skilled immigrants entered the labor market.³

Note that in these “perverse” cases, relative wage and employment effects have the opposite signs; for example, if wages of the unskilled (moderately) *increase* relative to wages of the skilled although migration was relatively unskilled, then it must be the case that employment of the unskilled (strongly) *declines* relative to the skilled. This emphasizes the need to investigate wage and employment responses due to immigration *jointly*, as investigating the two outcomes in isolation may paint a very misleading picture of the overall labor market effects of immigration.⁴

Finally note that – in the case of two factors of production – such “perverse” effect will only be observable when capital is not perfectly elastic, i.e. $\varphi < 0$. This is because in the case of perfectly elastic capital supply, the aggregate wage effect of a migration-induced supply shock is zero, and wage decreases for the skill group that receives a higher share of migrants and increases for the other skill group – regardless of the relative magnitude of the group-specific local labor supply elasticities. However, if we add a further skill group, then again such “perverse” effects are possible even if the capital supply is fully elastic (see Appendix B).

5.2.2 Heterogeneous Labor Supply by Subgroups

It is straightforward to extend the model to allow for heterogeneous labor supply responses across subgroups s (like for instance groups with different labor market experience)

³ The relative employment effect equals

$$dL_S^N - dL_U^N = \frac{(\eta_S - \eta_U)(\beta - 1)\Pi\left(1 - \frac{\varphi}{\beta - 1}\right) + (\beta - 1)\left[\eta_S \frac{\pi_S^I}{\pi_S^N}(1 - \varphi \eta_U) - \eta_U \frac{\pi_U^I}{\pi_U^N}(1 - \varphi \eta_S)\right]}{1 - (\beta - 1)[\eta_S(1 + s_S \varphi) + \eta_U(1 + s_U \varphi) - \eta_U \eta_S(1 + \varphi)(\beta - 1)]} dI.$$
 If η_S is large relative to η_U , employment of the skilled may decline relative to employment of the unskilled even if immigration is relatively unskilled.

⁴ Borjas (1999) also emphasizes that changes in the wage structure and in native employment need to be jointly considered to capture the full labor market impact of immigration. Our model implies that isolated estimates of wage or employment effects may not only understate the overall impact of immigration but also misrepresent its distributionary effects.

within a skill group g , so long as these are perfect substitutes in production.⁵ Assume that the local native supply response in group g in subgroup s , depends on local wage changes according to the subgroup-specific labor supply elasticity η_{gs} , such that

$$d\log L_g^N = \sum_s \eta_{gs} \frac{L_{gs}^N}{L_g^N} d\log w_g \quad (12)$$

The group-specific local labor market elasticity η_g may now be thought of as a weighted average of the subgroup-specific elasticities, where the weights equal the employment shares of each subgroup within a group. Since different types of workers within a skill group are perfect substitutes in production, the impact of an overall immigration shock on their wages is the same, and is given by equation (8). However, employment decreases most for the subgroup with the largest labor supply elasticity:

$$\begin{aligned} d\log L_{gs}^N \\ = \eta_{gs} \frac{(\beta - 1) \left[\frac{\pi_g^I}{\pi_g^N} (1 - \varphi \eta_{g'}) - \Pi(1 - \frac{\varphi}{\beta-1}) \right]}{1 - (\beta - 1) [\eta_g(1 + s_g \phi) + \eta_{g'}(1 + s_{g'} \phi) - \eta_g \eta_{g'}(1 + \phi)(\beta - 1)]} dI \end{aligned} \quad (13)$$

In the empirical analysis, we therefore investigate employment (and wage) responses across subgroups. The arguments here also highlight that the group-specific local labor supply elasticity η_g may be large even if some workers do not respond at all to local wage changes – so long as there are other workers who respond very strongly to such changes. Moreover, inelastic subgroups benefit from the presence of elastic subgroups, as the latter reduce, due to their strong employment response, the pressure on group-specific wages. This illustrates the way in which highly elastic subgroups “shield” less elastic subgroups from the impact of a labour supply shock.

5.2.3 Margins of Adjustment and the Local Labor Supply Elasticity

The analysis so far illustrates how the local labor supply elasticity and the variation of that elasticity across groups of workers are central determinants of the effects of immigration on local labor markets. We now illustrate that the overall response in local employment levels may stem from different margins of adjustment, which affect different type of workers and have disparate implications for the overall effect of immigration on the economy. The

⁵ Allowing for more than two groups with varying labor supply elasticities which are imperfect substitutes in production severely complicates the expressions for equilibrium wage and employment responses without leading to novel insights.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

discussion sheds light on the determinants of the local labor supply elasticity, and why this elasticity may vary across subgroups.

The local labor supply of natives in skill group g in some time period t (L_{gt}^N) can be expressed as a function of the native supply in the previous period (L_{gt-1}^N), inflows into local employment of natives who previously were not employed in the area (L_{gt}^I), and outflows from employment of natives who were locally employed (L_{gt}^O),

$$L_{gt}^N = L_{gt-1}^N + L_{gt}^I - L_{gt}^O \quad (14)$$

Native employment in an affected area may fall either because native workers leave local employment (*outflows*), or because less native workers enter local employment (*inflows*) in response to immigration. Its impact may thus fall on incumbent workers who are employed in an affected area – or on outsiders, who may want to enter the local workforce.

Within each of these categories we can further distinguish between movements from or into employment in other areas (*geographical* inflows $L_{gt}^{I,E}$ and outflows $L_{gt}^{O,E}$), and movements out or into the labor force (inflows $L_{gt}^{I,NE}$ from and outflows $L_{gt}^{O,NE}$ into *non-employment*). The percentage growth in employment between the two time periods can thus be decomposed as

$$\frac{L_{gt}^N - L_{gt-1}^N}{L_{gt-1}^N} = \underbrace{\frac{L_{gt}^{I,E}}{L_{gt-1}^N} - \frac{L_{gt}^{O,E}}{L_{gt-1}^N}}_{\text{employment}} + \underbrace{\frac{L_{gt}^{I,NE}}{L_{gt-1}^N} - \frac{L_{gt}^{O,NE}}{L_{gt-1}^N}}_{\text{non-employment}}. \quad (16)$$

The relative magnitude of these four margins is particularly important for the spatial approach, which exploits variation across areas to identify the effect of migration on native outcomes. The displacement of workers in affected areas may affect geographical flows and in turn employment in non-affected areas, thus contaminating counterfactual observations.

Moreover, the various channels of adjustment have different implications for the effect of immigration on the overall economy. Movements in and from employment in other areas affects the spatial distribution but not the level of employment in the national labor market, while movements in and out of the labor force have potentially implications for the overall level of economic activity.

Finally, each of the four margins affects different types of workers, potentially leading to disparate welfare implications, but also explaining why some groups may respond more elastically to wage changes than others. For example, older workers may be geographically

less mobile but also have access to better social security and unemployment benefits than young workers, and may thus be more likely to leave the labor force in response to adverse shocks. Skilled workers may be geographically more mobile than unskilled workers, and may be more likely to avoid localized shocks through spatial movement across labor markets.

To understand the consequences of immigration not only for aggregate outcomes in the affected area, but also for different types of workers and for the national economy, it is thus important to identify each of the four margins in equation (16). The nature of our data allows us to follow each worker over time, and across labor markets, and thus to estimate the relative importance of each type of employment adjustment.

5.3 Background and Data

5.3.1 The Policy

Our analysis makes use of a unique experiment that took place in a part of Germany that bordered the Czech Republic.⁶ In 1991, due to a sudden and unexpected policy change, Czech citizens were allowed access to the German labor market in a specific region along the Czech-German border. The policy, enabled only by the fall of the iron curtain in the preceding year, was part of a larger scheme for legal employment of foreign nationals in Germany. Introduced with effect to January 1st, 1991, the intention of the scheme's various provisions were to facilitate the recruitment of foreign workers, in a time of increased labor demand after German reunification.⁷ The particular provision we investigate is a commuting scheme ("*Grenzgängerregelung*"), which granted Czech nationals the right to work in dependent employment in German border regions, but did not grant residency, so that Czech workers were required to commute each day. The policy was otherwise non-restrictive, and was not constrained to specific industries or applicants with specific qualifications. Appendix C provides additional details.

The scheme applied also to Germany's second Eastern neighbor, Poland. However, as social security records in former East Germany are fully available only from 1992, and since its formerly state-directed economy was subject to strong structural change after German

⁶ The territory of the Czech Republic was part of Czechoslovakia until dissolved peacefully with effect to January 1st, 1993. For simplicity we use the term Czech Republic even when referring to the period before the separation.

⁷ See "Anwerbestoppausnahme-Verordnung", Bundesgesetzblatt, Jahrgang 1990, Teil I.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

reunification, we focus on the scheme's consequences along the Czech border with West Germany. Figure 1 maps the region that was affected by the scheme, comprising 21 districts that lie within an approximate eighty kilometers band from the Czech-German border. However, some of these districts are close to the former inner-German border and were thus after reunification in 1990 affected by commuters from East Germany, where wages were much lower than in West Germany. To avoid any contamination of our experiment, we exclude districts located within approximately 80 kilometers (our results remain robust to less conservative choices) of the inner-German border. As shown in Figure 1, this leaves us with a region that comprises 13 districts, or 291 municipalities. We refer to it below as the “*border region*”. The region is rural and contains various small, but no large cities. The first column of Table 1 illustrates that the local labor market had a comparatively small share of high-skilled workers with university degree, a young workforce, low wages, and a particularly low share of pre-existing immigrants.⁸

The introduction of the commuting scheme in January 1991 led to a substantial and rapid inflow of Czech workers into the border region. After 18 months into the scheme, a total of 9,996 workers from the Czech Republic were employed in the eligible districts. While negligible for the national economy, the inflow was large compared to existing employment in the border region. In Figure 2 we plot the employment shares of Czech nationals in border and selected inland districts (which we define precisely in the next section). By June 1992 the share of Czech nationals in the border region had increased from close to zero to about three percent, and on average to about 10 percent in municipalities that were closest to the border. In contrast, the employment share of Czech nationals in the inland, which was not affected by the commuting scheme, remained negligible.

The stark inflow was related to the large wage differentials along the border, as in the mid-1990s wages in Germany were multiple times higher than in the Czech Republic.⁹ The resulting labor supply shock was, at least prior to 1990, unexpected. It was also exogenous to the economic situation in the border region, as follows from the setup of the policy, which

⁸ See also Moritz (2011). This study differs from ours in that it is based on a 2 percent sample, in that areas with a high share of commuters from former East Germany are included in the analysis, and in that variation within the border region is not directly exploited.

⁹ See OECD data on average annual wages,
https://stats.oecd.org/Index.aspx?DataSetCode=AV_AN_WAGE#

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

applied non-discriminatorily along the border along all of the German borders with both Poland and the Czech Republic.

Figure 2 illustrates that the share of Czech workers remained stable in 1993 and decreased thereafter. Various reasons may have contributed to this decrease. First, Germany entered a recession in 1993, and the difference between German and Czech wage levels decreased steadily in the early 1990s. Both factors may have decreased the returns of commuting for Czech workers. Second, the large inflow of foreign workers caused a political backlash from late 1992. Allegations that foreign workers led to a worsening of conditions for native workers, while not leading to formal changes to the commuting policy, led to stricter local implementation of general restrictions regarding foreign employment in Germany.¹⁰

The fall of the iron curtain may have affected the border region also via other mechanisms, for example through cross-border demand for consumer goods, services and trade. While we find evidence for systematic structural changes in certain local industries, those responses are spread over a long time interval, in contrast to the sudden effects from the labor supply shock.¹¹ Their magnitude in the short run thus appear negligible compared to the extraordinary size of the supply-shock caused by Czech commuters. We further observe that the employment share in the retail sector remains relatively stable, and that employment and wage responses are clustered in those industries that received a larger share of Czech workers.

In the empirical analysis we focus on the immediate wage and employment effects of the labor supply shock up until 1993. The “reverse experiment” of the subsequent decline in the share of Czech nationals in 1994 and 1995 is interesting, but potentially endogenous to local labor market conditions. We abstain from the analysis of long-run outcomes as the spatial distribution of economic activity in Germany may shift in the process of further European integration.¹²

5.3.2 The Data

¹⁰ See Nürnberger Nachrichten, 1.7.1992, “*Konkurrenz auf Nordbayerns Arbeitsmarkt verschärft sich*”.

¹¹ An interesting case is the glass and ceramic industry, which for historical reasons was strongly concentrated on both the German and Czech side (*Bohemian glass* is the region’s famous export product since the Renaissance) of the border. After its opening we observe a sharp trend break in industry growth on the German side, but also find the subsequent decline of employment to be gradual, spread out over a full decade.

¹² For example, Redding and Sturm (2008) provide evidence that changes in market access caused by German division and reunification had a long-run impact on local population growth along the border.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

The data used in our analysis come from German Social Security Records covering more than two decades, from 1980 to 2001. They comprise the population of all men and women covered by the social security system; not included are civil servants, the self-employed, and military personnel.¹³ These data are uniquely suited for our analysis. First, the large sample size allows us to obtain fairly precise estimates of immigration on wages and employment even for detailed subgroups, although only a relatively small local area is affected by the inflow of immigrants. Second, the longitudinal nature of our data allows us to investigate whether the employment effects are driven by an increased outflow of workers into other area or into non- or unemployment, or by a decreased inflow of workers into the local labor market – a dynamic so far unexplored in the literature. Third, measurement error in the migration-induced supply shock, which, as illustrated by Aydemir and Borjas (2011), attenuates the impact of immigration on native labor market outcomes, is close to zero in our data. Finally, besides information on education, age and other individual characteristics, we also observe the citizenship of every individual that is in employment, which allows us to identify Czech workers that work in Germany (but live in the Czech Republic).

Our data is constructed so that we observe each individual as of June 30th of each year. Thus, the employment status of each individual refers to this date. In contrast, the wage variable records the average daily wage in the employment spell that contains the reference date.¹⁴ As it is typically the case with social security data, our wage variable is right-censored at the social security limit. In our sample, this affects about 3% of all observations. Following Dustmann et al. (2009), we have imputed censored wages under the assumption that the error term is normally distributed, and allowing for a different residual variance by age and education group as well as by district.

We distinguish two skill groups, which we label *unskilled* and *skilled*. The unskilled are workers who enter the labor market without postsecondary education. The skilled comprise workers who hold a high school degree, completed an apprenticeship scheme or equivalent, or graduated from a university or college.¹⁵ This classification of skill groups is meaningful in the

¹³ In 2001, 77.2 percent of all workers in the German economy were covered by social security and are hence recorded in the data (Bundesagentur für Arbeit, 2004).

¹⁴ As employers are required to update records only at the end of each year, the variable may capture wage changes that occurred after June 30th, up to December of the same year.

¹⁵ To improve the consistency of the education variable in our data, we impute missing values using the procedure proposed by Fitzenberger, Osikominu, and Völter (2006), using past and future values of the education variable. The imputed education variable is missing for 3.9% of observations in

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

German context, where many apprenticeship jobs educate for professions that require college degrees in Anglo-Saxon countries (such as medical assistant or bank clerk). We do not report separate results for university or college graduates, as their share in the border region in 1990 is less than 5%. Within each of these skill groups, we also distinguish three age groups: younger than 30, 30 to 49, and 50 and older.

We restrict the analysis to individuals aged between 18 and 65 and we exclude irregular, marginal and seasonal employment as well as individuals in apprenticeship training. Our analysis of employment effects is based on workers in regular full- and part-time employment, where we down-weight those in part-time employment into full-time equivalent units by 0.67 (“long” part-time) or 0.5 (“short” part-time). Our wage analysis is based on full-time employed workers only.

Table 2 provides descriptive statistics on both the existing stock of workers in the border region in 1989 (before workers from the Czech Republic entered), and of Czech nationals, whose characteristics we measure in 1992. Czech workers are far more likely to be unskilled than the existing workforce (50.5% vs 27.6%).¹⁶ As the education information stems from German employers, our classification is less likely to reflect their formal level of qualification in their home country but their perceived level on the German labor market. Imperfect transferability of country-specific qualifications are thus less likely to bias classification than in studies that rely on individual surveys. Moreover, Czechs commuting into the border region are more likely to fall into the age group 30 to 49 than natives (61.9% vs 40.8%), whilst the shares of workers older than 50 are much lower among Czechs (3.7% vs 15.7%). Czech nationals are predominantly male, often speak German, and on average earn much lower wages than natives.¹⁷ They are over-represented in construction, the hotel and restaurant industry, wood processing and manufacturing, and under-represented in the public sector.

the overall sample, and 2% of observations in the border region. We classify these individuals as unskilled, but this choice has little impact on our findings.

¹⁶ In Czechoslovakia, about 15 per cent of the age group entered *gymnasia* in 1989 (mostly as a preparation for university studies); about 25 per cent attended technical (occupationally specialized) education; and 60 per cent entered apprenticeships (often with subsequent entry into employment in the enterprise where the training had been conducted). Source: OECD Country Note on the Czech Republic, November 1997.

¹⁷ Our data does not contain information on language proficiency, but a high proficiency of German is suggested by the geographic proximity and confirmed by reports in the press (see for example Nürnberger Nachrichten, 1.7.1992).

5.4 Empirical Strategy

In this section we discuss our empirical strategy in details. Section 4.1 explains how our main regression equations relate to the theoretical model that we presented in Section 2. In the following three sections we provide details on conceptual issues that are either not modelled within our theoretical framework or that are specific to our empirical implementation, such as the spatial distribution of Czech inflows (section 4.2), the choice of control units (section 4.3), the timing of native responses, and the potential for specification or falsification tests that are possible in our data (section 4.4).

5.4.1 The Effect of Immigration on Wages and Employment

Our basic estimation equation regresses the change in log wages of German nationals (*natives*) in skill group gs and area j between two periods ($\Delta \ln w_{gs,j}$), or their percentage change in employment ($\Delta N_{gs,j}$), on the total inflow of Czech workers between 1990 and 1992 as a share of total employment in that area in 1990 (ΔC_j),¹⁸

$$\Delta \ln w_{gs,j} = \alpha_{gs} + \beta_{gs} \Delta C_j + u_{gs,j} \quad (17)$$

and

$$\Delta N_{gs,j} = \gamma_{gs} + \delta_{gs} \Delta C_j + v_{gs,j} \quad (18)$$

The parameters α_{gs} and γ_{gs} measure skill group-specific growth rates in wages and employment. The parameter β_{gs} measures the impact of an inflow of Czech workers between 1990 and 1992 on the (change in) wages of native workers in skill group gs between two time periods, whose choice we discuss in section 4.3 below. This parameter corresponds to the expression we have derived in equation (10) above. Hence, its sign and magnitude depend on the relative density of immigrants to natives in the skill cell g , as well as on the elasticities of capital supply, and the own and the other groups labor supply elasticities. The parameter δ_{gs} in equation (18) corresponds to the expression in equation (13). It follows that the ratio δ_{gs}/β_{gs} identifies the labor supply elasticity of skill group gs , η_{gs} .¹⁹

¹⁸ As we point out above, the main inflow of Czechs happens in 1991 and 1992; since the share of Czech workers in 1993 is almost the same as in 1992, adding an additional year would hardly affect our regressor.

¹⁹ Note that even if, within education groups, age groups are imperfect substitutes in production, dividing the subgroup-specific employment effect by the corresponding wage effect identifies the subgroup-specific local labor supply elasticity.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

Note that this specification does not require us to pre-allocate *immigrants* to particular skill cells. For each skill cell that we consider for natives, the regressor will always be the overall inflow of immigrants; the size and sign of the estimated parameter will depend on the (elasticity-weighted) relative density of immigrants in the particular native skill group that we consider (see Section 2). The estimated parameter has thus a clear interpretation. This specification avoids not only the problem of misclassification, which arises if immigrants are assigned based on their observed characteristics into skill groups within which they do not compete with natives.²⁰ It is also consistent with our experiment, as only the *total* inflow of Czechs into the border region can be considered as quasi-random. In contrast, how many Czechs from a particular education or age group enter may well be endogenous, and could be driven by the specific conditions in the region. Finally, the estimated parameter is policy relevant, as it captures the total effect of the aggregate supply shock, for specific groups of natives. In contrast, parts of the literature identify only relative or distributionary effects between groups, or relate their responses only to skill-specific shocks.

5.4.2 Exploiting Distance to Border

As described in section 3.1, the sudden and unexpected introduction of the commuting policy can be seen as a quasi-natural experiment. A comparison of the border region that was affected by the policy with non-eligible control areas can thus solve the endogeneity problem that immigrants may self-select into areas in which the economy is growing.

However, the unique commuting-nature of the experiment provides additional variation in the exposure of different areas to Czech inflows that can be exploited. Czech workers were not allowed to live in the border region but instead had to commute on a daily basis.²¹ As commuting is increasingly costly the farther away an area is located, municipalities close to the border were more attractive and potentially more exposed to the policy.

Figure 3 illustrates that distance to border is indeed a key determinant of where Czech nationals locate within the border region. In the figure we plot, for municipalities within the

²⁰ Dustmann and Preston (2012) illustrate that assignment of immigrants to skill groups due to observed characteristics may lead to serious misclassification, as immigrants often *downgrade* upon arrival.

²¹ The requirement to commute was enforced via various channels. First, workers that entered employment under the commuting scheme had to apply for a special type of permit, the “*Grenzgängerkarte*”, revealing the worker’s conditional residence status. Second, in Germany it is compulsory for residents to register with the local registry office. Importantly, double registration was required, as both tenants and landlords were obliged to submit information.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

border region, Czech employment growth from 1990 to 1992, ΔC_j , against the municipality's distance to the closest border crossing, Z_j , where we weight municipalities according to their employment in 1990. The figure shows that municipalities next to the border received the largest inflow of Czech workers, corresponding on average to almost 10% of employment in 1990. In contrast, municipalities located more than 50 kilometers away from the border experienced hardly any inflow. We exploit this spatial variation for identification by using distance to border as an instrumental variable (IV).

We report the corresponding regression results (the first stage) in Table 3, where we approximate the relationship between the inflow of Czech nationals and distance to border as a quadratic function. The coefficients on distance and on distance squared are jointly highly significant, and together explain 38.7% of the variation in the Czech employment share across municipalities within the eligible border region, or 54.4% of the variation across border and matched inland municipalities (see below). The rank condition is thus satisfied. The exclusion restriction requires that distance to the border affects group-specific labor market outcomes only through its relation to the inflow of Czech commuters. This assumption may be violated if areas closer to the border experience different economic trends, or if the causal effect of Czech commuters on heavily exposed areas lead to spatial spillover to adjacent areas. Below we address both, by estimating placebo tests in pre-experiment periods, by directly estimating the magnitude of spillovers between adjacent areas, and by comparing our baseline results to other specifications that rely on a different source of variation.

5.4.3 Selecting Control Units

From the previous section it follows that we can estimate equations (17) and (18) in three different ways. First, by comparing the region that was eligible under the commuting policy against suitable control areas that were similar in observable characteristics, but not eligible. We describe below how such areas can be selected. Second, by exploiting only (instrumented) variation in the exposure to Czech inflows across municipalities within the affected border region. Third, we can combine the two approaches, pooling municipalities within the border region with inland areas that were not exposed. We choose this last approach as it exploits the available variation more thoroughly, leading to more precise estimates and allowing us to analyze the response among natives in more details.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

However, as a robustness test we also report separate results based on the first and second approach. Using only variation across areas within the border region has the advantage that those areas are very similar in terms of spatial location, education, age, and industry composition, and are thus likely to experience similar economic time trends. However, areas that are less affected by immigration may nevertheless be subject to spillover effects from areas that are more affected, thus contaminating counterfactual observations. For instance, if natives from areas close to the border search for jobs in adjacent areas in response to the labor supply shock we may overstate the employment response to migration. We assess the importance of such “spillovers” directly below and find that they are small. Alternatively, the problem can be addressed by comparing the inner border region with a set of control areas from the inland, which are located sufficiently far away from the German-Czech border. These areas are not affected by spillover or general equilibrium effects, as the regional supply shock was small in national terms. We implement such comparison also by using a synthetic cohort method, following Abadie et al. (2010). We find that all three approaches lead to very similar results.

For the matching of inland control units we consider only West-German districts of similar urban density (rural areas or areas with intermediate agglomerations). We also exclude districts that are within eighty kilometers from the former inner-German border, to avoid contamination from German reunification. We match then on variance-weighted differences in the employment share of three education groups, the employment share of foreign nationals, mean log wages, the share of right-censored wage observations, local employment levels, and the employment shares of four age groups in June 1989 (the last observation before reunification and the fall of the iron curtain). To account for differences in the size of districts we match, for each treated district, one or multiple control unit(s), until their employment levels sum to at least proportion x of employment in the treated unit. The choice of x is subject to a trade-off between bias and precision; choosing a higher value results in the matching of more but potentially less suitable control areas. We choose $x = 1.5$ in our baseline. To ensure that our findings are not driven by the particular choice of control units we repeated our analysis for alternative sets of match characteristics and other values of x .

Figure 1 shows the 24 matched inland districts. In the three columns of Table 1 we compare border districts, all other areas, and matched inland districts in terms of observable

characteristics in June 1989. Compared to all other inland districts, border districts have a somewhat less educated and younger workforce. Median wages are considerably lower in the border region (about 15%), as is the share of foreign nationals. As expected, matched inland districts are more similar to border districts, including in their industry structure.²² Interestingly, Figure 2 shows that matched control areas tend to be likewise located along Germany's borders, reflecting the observation that remote and border areas tend to be characterized by lower economic activity, smaller firms, and lower wages.

5.4.4 Timing and Placebo Tests

At this stage it is unclear over what time period the dependent variable should be observed, as our model – as is the case for most theoretical work in the literature – does not provide implications about the precise timing of events. The choice of period in empirical work is instead often predetermined by data constraints (e.g. 5 or 10 year time intervals), or the dependent variable is simply defined over the same period as the immigrant inflow. However, the response of wages and employment to immigration-induced supply shocks may not be instantaneous, and very little is known about its respective evolution over time.

We adapt instead a flexible specification that allows us to assess the timing of events. Our regressor ΔC_j is always defined as the inflow of Czech workers between 1990 and 1992 into area j as a share of local employment, but we specify the dependent variable over different time windows. In particular we obtain coefficients from annual regressions – regressing wage or employment changes between the years t and $t-1$ on ΔC_j . We weight each regression by the employment of (sub)group gs in area j in year $t-1$. To obtain the overall impact of the labor supply shock over longer periods we then add up the respective coefficient estimates.

As a placebo test, we estimate equations (17) and (18) also for years before 1990. Obviously, the inflow of Czech nationals into an area between 1990 and 1992 should have no impact on native employment changes prior to 1990. Hence, the hypothesis $H_0: \beta_{gs} = 0$ and $H_0: \delta_{gs} = 0$ for $t \leq 1990$ provide us with interesting falsification or placebo tests, against which we can test the identifying assumption that areas that are located close to the border

²² The exception is the glass and ceramic industry, which for historical reasons is heavily concentrated in the border region.

experienced the same time trends prior to 1990 as areas located further away.²³ Coefficients for $t > 1990$, in contrast, identify the causal impact of immigration, and are informative about how fast employment and wages respond to the inflow of Czech workers.

Running yearly instead of “long” difference regressions also allows us to address potential selectivity bias in wage estimates. Empirical studies typically compare the average wage of all workers employed in a given area and time period to the average wage of all workers who are employed in another period. However, as pointed out by Bratsberg et al. (2012), such estimates of the wage effect can be severely affected by worker selection, i.e., by high or low ability workers selectively moving into and out of the area (or the labor market) in response to migration. To address this issue we assess wage effects only for those “incumbent” workers who were employed in the same area in *both* t and $t-1$.

5.5 Results

5.5.1 Main Results

Figures 4a and 4b show the cumulative effect of the inflow of Czech commuters on the overall employment and wages of native workers over the period between 1985 and 1995. The figures are based on the estimation of equations (17) and (18), where we regress the change in native wages or employment between two consecutive years on the inflow of Czech workers between 1990 and 1992, instrumented by the area’s distance to the border.²⁴ We then plot the cumulative effects, starting in 1990, by adding up estimated coefficients backwards and forwards.²⁵ Thus, the figures represent the cumulative wage (or employment) effects of an inflow of Czech workers between 1990 and 1992 for each year between 1985 and 1995.

As we should expect the Czech inflows between 1990 and 1992 to affect neither wages nor employment in the years before 1990, the figure entries before that date serve as a placebo or “falsification tests” for the comparability of our treatment and control areas. The estimated

²³ Falsification tests have been used in various literatures, see Abadie et al. (2010). Within the migration literature, Angrist and Krueger (1999) perform such test to show that the estimated effect of the famous Mariel Boatlift on the Miami labor market is sensitive to differences in trend between treatment and control units.

²⁴ We pool municipalities from the border and inland regions and weight by the number of native workers employed in that particular municipality (see Section 4). We instrument the inflow of Czech nationals with an indicator variable equal to 1 if the area is part of the border region (and 0 if it belongs to the inland), and add an interaction of this indicator variable with distance to border crossings and distance to border crossings squared. The inflow of Czech immigrants to inland areas is basically zero.

²⁵ Standard errors are computed by the bootstrap method and take the covariance of year-specific coefficient estimates into account.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

coefficients are indeed small and statistically not significant. The employment coefficients are clustered closely around zero, illustrating that native employment grew by a very similar magnitude in treatment and control areas. The wage development was slightly more positive in areas close to the border. Below we also present results that control for such differences in pre-treatment trajectories.²⁶

After 1990, however, the figures show that wages and in particular employment of native workers drop significantly. For instance, Figures 4a and 4b suggest that by June 1992, a one percent inflow of Czech workers has led to a wage decrease of natives in the border region by about 0.12 percent, and a decrease in employment by about 0.6 percent (relative to 1990). Given that the average share of Czech workers in the affected border region in 1992 is about three percent (see Figure 2), average wages have decreased by about 0.36 percent in the region (3×0.12), and native employment by about 1.8 percent (0.6×3). The average share of Czech workers is about ten percent in municipalities closest to the border, in which average wages and employment decreased thus by 1.2 and 6 percent, respectively.

While the average wage effect is relatively modest, the employment effect is quite substantial. Below we investigate in more detail the composition of this effect, across groups of workers and different margins of adjustment. Here we would like to note that average wage effects – within the simple model we have pointed out above – imply that the supply of capital is not fully elastic in the short run.

Figures 4a and 5b further suggest that the growing number of Czech commuters, which reached its peak level in mid-1992, continued to have adverse effects and depressed native wage and employment growth until mid-1993. In Table 4, we provide cumulative estimates within this time period that are broken down by skill group. In row (i) we present the effect on wage growth between 1990 and 1991 of an inflow of Czech workers for the same period. Rows (ii) and (iii) present estimates for the cumulative effect on wage growth between 1990 and 1992 or 1993 of an inflow of Czech workers between 1990 and 1992. Finally, row (iv) presents the same specification as row (iii), but reports the OLS coefficient instead. Note that the skill group-specific specifications in Panels B and C only differ in the dependent variables, but the regressor is always the overall inflow of Czech workers between 1990 and 1992

²⁶ Note that we selected control units based on the *level* of wages and employment in 1989, not on their pre-treatment trajectories. The finding that those trajectories are nevertheless similar in treatment and control areas is thus important.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

(except for row (i), where it is the inflow between 1990 and 1991). The estimates correspond therefore to the expression in equation (10), and measure the overall impact of Czech workers on wage or employment growth of native workers in the respective skill group. In the last column, we report the labor supply elasticities, which we obtain by dividing the coefficient estimates in the second column by estimates in the third column.

Results in Panel A suggest that the inflow of Czech workers had a sizeable and rapid impact on both wages and employment of native workers. The wage effect has been quite immediate and relatively constant, while native employment responds also in 1991, but more than doubles until 1993. The comparatively slower response in employment leads to an increase in the estimated labor supply elasticities, from 2.6 to nearly 7 between 1991 and 1993. These results illustrate that the degree to which the effect of a local supply shock is felt in either wages or employment depends crucially on the time frame under consideration. While wages respond more quickly, we find that the immigration-induced supply shock has also immediate effects on native employment, which has important implications for studies that examine the wage effects of immigration.²⁷ We explore the mechanisms that permit such rapid adjustments of aggregate employment below.

Inspection of Panels B and C shows that the response of unskilled workers to the inflow of Czech commuters has been larger than the response of skilled workers, both in terms of wage and employment responses. For instance, and considering the overall impact over the period between 1990 and 1993, a one percent point increase in Czech workers has decreased wages and employment for unskilled natives by 0.2 and 1.4 percent, while it has decreased wages and employment of skilled natives by 0.1 and 0.5 percent, respectively. Given that some areas experienced an inflow of Czech workers corresponding to 10 or more of percent local employment, in particular the employment effects are substantial. The stronger decline of wages and employment among unskilled natives is expected from our theoretical model, as Czech commuters were mostly unskilled (see Table 2). However, our results suggest that the supply shock had adverse effects in both native skill groups. This is important, as parts of the literature consider only distributionary effects between skill groups, assuming that the overall effect of immigration on native outcomes is zero. Rows (iv) reports estimates when we do not

²⁷ For example, Borjas (1999) notes that wage changes in isolation are informative about structural parameters only if measured before responses in native employment occur. Our results suggest that the observation of wage changes are unlikely to suffice even if measured right after an immigrant-induced supply shock occurs, since native employment responds too rapidly.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

instrument the share of Czech workers by distance to the border, using the specifications in row (iii). The Table entries show that the estimated wage- and employment effects are substantially smaller than the IV estimates, which is to be expected if Czech workers entered predominantly those municipalities that experienced higher wage growth. Interestingly, the difference between OLS and IV estimates is more striking among the skilled, suggesting that selection is a larger problem for skilled than for unskilled workers.

Panels B and C show that not only the magnitude, but also the speed by which skilled and unskilled natives are affected differs strongly. The employment response among unskilled natives is nearly immediate, while it evolves more gradually among the skilled. As a consequence, the estimated labor supply elasticity is nearly constant among unskilled but increases for skilled natives.

5.5.2 Robustness Checks

In Table 5, we provide different sets of robustness checks. The first column in the Table reports, as a reference point, the estimates from row (iii) in Table 4. In column 2, we report trend-adjusted estimates, where we identify the municipality-specific trend in wage growth in each municipality from information for the years 1987-1989. Estimates are quite similar as those we report in column 1. In column 3 we report long differences. These estimates are obtained by regressing wage or employment differences between 1990 and 1993 on the inflow of Czech workers between 1990 and 1992, but (log) wages are averaged over all workers who are in employment in each those two years. Note that this is what is usually done in studies that are based on (repeated) cross-sectional data. It differs from our estimates in the other columns, where we add up yearly wage effects that are averaged only for those workers who remain employed in two consecutive years. The key difference is that long difference estimation does not consider that the composition of the workforce may change as a result of immigration. Given the magnitude of our employment effects, this is likely to induce some serious selection. As it is lower productivity workers who are most likely to be affected by immigration, this will lead to a bias towards zero in the wage effects.

Estimates of employment effects are hardly affected and very similar to those in the first column. This illustrates that for employment effects, our strategy to aggregate year-specific regression coefficients, while more flexible, does not lead to distortions in comparison to the

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

direct estimation of long differences. However, the way wages are measured is very important. Based on long difference estimations, the results point at no overall wage effects of the inflow of Czech workers. The group-specific estimates for skilled and unskilled workers are still negative but smaller in magnitude than in our baseline, more so for the unskilled. These results are due to compositional changes within and between skill groups; within each skill group, workers with low wages are more likely to leave or to not enter the workforce in response to migration. A simple comparison of average wages before and after the migration-induced supply shock thus underestimates the wage effect on the remaining workers. Since the loss of native employment is larger among the unskilled, who tend to have lower wages, the overall wage estimates are more biased than the skill-specific wage effects.

Thus, comparison of results in column 1 and column 3 suggest that, if immigration leads to selective employment effects – as is expected from a theoretical model as described in Section 2 – estimation based on repeated cross sections that are some years apart may not be able to pick up wage effects. Based on the estimates in column 5 only, we would probably have wrongly concluded that overall wage effects are close to zero, and skill-group specific wage effects are very small and not precisely estimated.

In columns 4 and 5 we use alternative sets of control units in the regressions. In column 4 we compare only the highly affected Eastern (“inner”) part of the border region, which due to its shorter distance to the border received the vast majority of Czech inflows, to unaffected areas from the inland.²⁸ In column 5 we compare only differentially exposed areas within the border region, dropping all control areas from the inland. As discussed in Section 4.3, these different set of controls have distinctive advantages and disadvantages. The results are nevertheless very similar to our baseline estimates in both cases, which is re-assuring.

Finally, we follow an alternative estimation strategy that exploits the commuting policy directly and does not make use of the variation in exposure across areas that was due to their different distance to the Czech-German border. We apply the synthetic control method proposed by Abadie et al. (2010), which compares a single treatment to a weighted average of available control units. We define our treatment unit as the heavily exposed inner border region, and construct a comparison unit from the matched inland districts according to the

²⁸ We split municipalities within the border region according to their fitted values from the first stage regression. The inner border region is comprised of 145 municipalities in which the predicted inflow of Czech was above the median, averaging to about 5.8 percent of total employment.

following procedure. First, we define the pre-intervention period as the years 1983 to 1990. Second, we define a vector \mathbf{X}_1 of pre-intervention characteristics for the exposed region that consists of the value of the outcome variable 1985 to 1989 and the average over the entire pre-intervention periods of employment growth, wage growth, the share of unskilled, the foreign share, and the share of four age groups. Similarly, define \mathbf{X}_0 to be a matrix that contains the same variables for the unaffected areas. Third, a weighting vector \mathbf{W}^* is chosen to minimize the distance $\|\mathbf{X}_1 - \mathbf{X}_0\mathbf{W}\|_V = \sqrt{(\mathbf{X}_1 - \mathbf{X}_0\mathbf{W})'V(\mathbf{X}_1 - \mathbf{X}_0\mathbf{W})}$, where V is chosen among positive definite and diagonal matrices such that the mean squared prediction error of the outcome variable is minimized for the pre-intervention periods.²⁹ The synthetic control method thus sets both a weight for each predictor (via V); and a weight for each available control district (via W).

As in our main analysis we consider the percentage growth in native employment and the difference in log mean wages of native incumbents as main outcomes, and consider accumulative growth, relative to the final pre-treatment year 1990. Figures 5a and 5b compares the evolution of aggregate native employment and wages in the treatment and synthetic control units. Pre-treatment trends are more similar than in our baseline specification since the choice of control units is now directly dependent on those trends. Native wages and employment in the treatment unit decrease substantially between 1990 and 1993 in comparison to the synthetic control. For comparison to our baseline specification, those differences in employment growth (-0.057 by 1993) and wages (-0.008) need to be scaled by the share of Czech workers that entered the treatment region (5.8 percent). The results, $-0.057/0.058=-0.978$ for employment and $-0.008/0.058=-0.136$ for wages are very close to our baseline coefficients. Differences between skill groups are likewise reproduced, with substantially larger employment and wage responses among unskilled natives.

5.5.3 Age-Group specific Responses

As we explain in Section 2, within the same skill group, overall labor supply elasticities may be large even if some workers do not react at all to a change in wages, as long as other

²⁹ See Abadie and Gardeazabal 2003, Appendix B. For implementation we use the software package “Synth”, which is provided by Abadie, Diamond and Hainmueller at <http://www.stanford.edu/~jhain/synthpage.html>.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

workers react strongly. The presence of highly elastic subgroups within the same skill groups “shields” other workers from a labor supply shock.

To investigate this further, we consider one individual characteristic that may relate to the elasticity of labor supply, which is age. Different age groups may react quite differently to a supply shock-induced change in wages, due to differences in their outside options, which generates different labor supply elasticities. More particularly, older workers may have more opportunities to withdraw from the labor market, for instance through early retirement, or by using an unemployment or disability spell to bridge the time to retirement.³⁰ On the other hand, such withdrawal may be much more costly for young workers, or workers at the peak of their career, who may not only lose earnings opportunities, but also experience a damaging loss of human capital.

In Table 7 we report results by age groups. Panel A reports estimates for all workers, while Panels B and C report results for unskilled and skilled workers. The first three columns report estimates where we do not adjust for pre-trends, while the last three columns report trend-adjusted estimates.

The estimates in Panel A suggest that the wage responses are largest for young and small for older workers, consistent with the hypothesis that workers of different age groups are imperfect substitutes and the relative density of Czech workers across age groups (Table 2). However, employment responses are *largest* for native workers in the oldest age group, even though only few Czech workers entered employment in the border region. Focusing on the trend adjusted results in the second set of columns, an immigration-induced labor supply shock by 3 percent as a share of total employment decreases employment among older workers by about 4.5 percentage points. In contrast, the employment response among younger workers, whose wages were substantially stronger affected, is less than half as large.

These results suggest that the elasticity of employment with response to wage changes varies substantially with age. We allowed for this possibility in our theoretical model in Section 2, and explained that such heterogeneity has important implications for the skill-cell approach in the immigration literature. The approach assumes that employment responses are proportional to wage responses across finely defined skill groups, such that differences in the

³⁰ Regulations with regard to early retirement and the prolonged provision of unemployment benefits were comparatively generous in Germany in the early 1990s. For example, the so-called “58er-Regelung” permitted workers who were at least 58 years old to leave the labor force while continuing to receive unemployment benefits until retirement.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

density of immigrants across those groups can be used for identification. Our findings show that this assumption may not always hold, as native responses can be largest in groups that receive a comparatively small share of immigrants.

Inspection of the labor supply elasticities shows a clear pattern, where elasticities are largest for the oldest age group (columns 4 and 7). These differences are also statistically significant: the final two rows in each panel shows test statistics for the hypothesis that the elasticity of older workers is greater than elasticity in the youngest age group, presenting the share of bootstrap samples (clustered on municipality level) for which this is the case.

Distinguishing workers by skill group leads to a similar pattern for skilled workers, while employment responses for the unskilled seem to be more uniform across the different age groups. Skill matters particularly for young workers – across all skill-age groups, the employment response was strongest among young unskilled, and particularly weak among young skilled workers. Estimates of the elasticity of native employment with respect to wage changes differ substantially with age also within skill groups.

5.5.4 Outflows versus Inflows

As discussed in Section 2.3, the total change in employment that we report consists of two key components: workers who leave employment in a particular area (“outflows”), and workers who do not enter employment, but who would have done so in the absence of a labor supply shock (“inflows”). These two movements together determine the employment effects that are usually estimated in the literature. Outflows and Inflows can then be further decomposed into outflows (inflows) into (out of) non-employment, and outflows (inflows) into (out of) other areas. The relative magnitudes of these different sources of aggregate employment changes are important to better understand the effects immigration-induced labor supply shocks have on the native workforce, and to gain further insight into the nature of variation in labor supply elasticities that we have discussed.

So far however very little is known about these underlying dynamics. We are in the fortunate position, due to the unique data that we are able to draw on that follows individual workers over time, and the nature of our underlying experiment, to assess the magnitude of these different channels of adjustment.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

Figure 5 provides a first illustration. The figure decomposes the overall (accumulated) change in native labor supply into responses due to changes in inflows (panel A) and outflows (panel B). Other than Figure 4, the graphs do not represent cumulative but yearly responses, which are estimated based on the same regressions than those used in Figure 4.³¹ An unusually high share of workers disappear from social security records in 1988 in various areas within the border region, many of whom reappear in 1989. A detailed analysis of these outliers suggests that the disappearance was likely triggered by a large military maneuver that took place in those areas at the time.³²

The figures indicate that overall yearly employment effects are mainly driven by a reduction in the inflow of new entrants in affected areas in the border region, and to a far lesser extent by an increase in outflows. Furthermore, while the inflow response is almost immediate, the outflow response is slightly delayed. This is not unexpected, as decisions not to enter a particular labor market are likely to be more responsive to local wage changes than decisions to leave that market. The immediate response of inflows can in turn explain why we find native employment levels to respond so rapidly to local shocks.

As we discuss earlier, inflows in and outflows out of employment can in turn consist of movements out of (and into) other regional labor markets, and movements out of (into) employment. Spatial movements are the reason why some have been critical about the area approach to identify the impact of immigration on wages (Borjas et al. 1996, Borjas et al. 1997) – as such movements dissipate its impact across areas, biasing the estimated effect towards zero. However, others have suggested that there is little evidence for spatial movements as a response to labor supply shocks (Card and DiNardo 2000, Card 2007). To investigate this in more detail, we cut our employment effects along various dimensions, distinguishing between movements from and to other areas, and movements from and to employment.

³¹ Inflows in group g are defined as $L_{j,gt}^I/L_{j,gt}^N$, where $L_{j,gt}^I$ is the number of native workers who are employed in municipality j in year t , but who were not employed in this municipality in the previous year. The definition of outflows is accordingly.

³² About 3000 members of the armed forces and many workers in potentially related establishments disappear temporarily in 1988. This observation is likely related to the annual “*Return of forces to Germany*” series of NATO military exercises, which took place in the Bavarian border region in this year. “*REFORGER 88: Certain Challenge*” was the largest European ground maneuver since the end of World War II (125,000 troops were deployed; see *The Stars and Stripes* (1988), Vol. 47, No. 147).

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

In Table 7, we report these decompositions for overall employment effects (Panel A), and separately by skill (Panel B) and age groups (Panel C). As in such fine categories the pre-treatment trends tend to differ across areas we report estimates that control for those differences using data from the years 1989 and 1990. As such, the overall employment estimates reported in the first column differ slightly from those reported in Tables 4 and 6. Entries in columns 2 and 3 are decompositions into geographic movements from and to employment in other areas, and movements from and to non-employment.³³ In column 4 we widen our definition of geographic movements by tracking individuals over longer time periods, counting also those individuals who come from (leave into) non-employment as geographic movers who, within a 3-year window, were employed in a different area before (after) their non-employment spell. Within that broader definition of geographic movements we distinguish between those who moved only short distances (into or from a different municipality within the same district) and those who moved further away. We report estimates of the immigration-induced supply shock on the latter category in column 5. Finally, in column 6 we report estimates on the response in population levels.³⁴ Population counts across municipalities and years come from the Federal Statistical Office in Germany, and constitute an alternative data source to assess the magnitude of geographic movements that occur in response to the inflow of Czech commuters. However, these counts are not separated by age or education.

The figures in columns 2 and 3 show that both channels, movements from and to employment in other areas, and movements from and to non-employment within the affected area, are important in explaining the overall employment effect. However, movements from and to non-employment are far more relevant in terms of magnitude than movements across area. Focusing on the overall impact on employment, numbers in Panel A suggest that, while an increase in the share of Czech workers between 1990 and 1993 by 1 percent decreases employment by native workers by nearly as much overall, about 17 percent of this adjustment

³³ Geographic inflows from employment in other areas in group g are defined as $L_{j,gt}^{I,E}/L_{j,gt}^N$, where $L_{j,gt}^{I,E}$ is the number of native workers employed in municipality j in year t who were employed in another municipality in the previous year. Inflows from non-employment are defined as $L_{j,gt}^{I,NE}/L_{j,gt}^N$, where $L_{j,gt}^{I,NE}$ is the number of native workers employed in municipality j in year t who were not employed in the previous year. Our data allow us to further distinguish between those individuals who were registered as unemployed from those who were not; corresponding estimates are available upon request. The definition of outflows is accordingly.

³⁴ The dependent variable in population regressions is the percentage growth in population levels between year t and the previous year.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

is due to workers either not entering the respective local labor market, or moving out of that labor market, while the remaining 83 percent are due to movements into and out of employment. Column 4 shows that while the relative importance of geographic movements increases when its definition includes movements through non-employment, the coefficient estimates are still far lower than those on movement into and out of non-employment within a local area. Column 5 suggests that about half of the employment reduction through spatial movements is due to changes in the number of movements over long distances.

Finally, the estimate in column 6 suggests that a 1 percent increase in the share of Czech commuters – who do not live in the affected German border region – decreases local population levels by 0.3 percent. This coefficient is close to the estimated geographic displacement of native workers that we estimated based on the tracking of individual workers in the social security data, reported in column 4. While the two coefficients do not have to match – workers who leave or do not enter exposed areas in response to Czech inflows may have more or less dependents than the average native workers – it is nevertheless reassuring that the evidence from a second, independent data source is consistent with estimates from our more evolved analysis of the social security data.

When splitting up these decompositions by skill groups (Panel B), the numbers show that for unskilled workers, almost the entire employment effect is due to movements into and out of employment within the local labor market, while for the skilled, geographic mobility plays a more important role: For these workers, almost 25 percent of the overall employment response is due to movements across areas.³⁵ This suggests that different skill groups adopt different responses to labor supply shocks, with the effect of immigration on skilled workers being more quickly dispersed across areas than the effect on unskilled workers.

In panel C, we distinguish between different age groups instead. Again, the first column reports the overall employment effects. The largest employment response is by individuals in the oldest age group, almost all of which is due to movements into and out of employment. In stark contrast, young workers (those below the age of 30) react far more strongly through changes in geographical movements, while workers in the age range between 30 and 49 have an intermediate position in this respect. These findings point at different underlying

³⁵ These findings are consistent with the general mobility patterns across skill groups. While only about 3.5 percent of all unskilled workers switch from employment in one to another municipality within a year, the share is close to 6.5 percent for skilled workers.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

mechanisms for overall employment responses of workers in different age- and skill groups to local labor supply shocks. While the unskilled and older workers absorb such shocks almost entirely by changes in movements into and out of employment within those areas that are affected by labor supply shocks, skilled and younger workers respond also by adjusting their movements across areas. Note however that the overall employment response of these latter groups is far smaller, which explains that overall – and as reported in Panel A – the response to local labor supply shocks by movements across areas is far less important than by movements from and to employment.

In Table 8, we break the different responses down further. The structure of the Table is the same than Table 7, with Panel A reporting the employment effects for the overall population, while Panels B and C distinguish by skill- and age groups. However, other than in the previous Table, we now decompose the total employment effect into inflows and outflows (column 1 and 2; these are the cumulated yearly effects between 1990 and 1993 that are displayed in Figure 6a and 6b). We then further decompose inflows and outflows into movements from (to) non-employment, and movements from (to) other areas (columns 3-6 and 5-8).

As suggested by Figure 5, the numbers in the first two columns in Panel A show that overall, inflows are far more important to explain the total employment response than outflows, a finding that stands in stark contrast to the way employment responses to immigration are commonly interpreted. Furthermore, columns 3 and 4 suggest that such inflows are mainly driven by a reduction of inflows from non-employment (73 percent of the overall effect), with the reminder driven by a reduction in inflows from other labor markets. Column 5 shows this finding to be robust to a widened definition of geographic inflows, counting also those individuals who come from non-employment but were previously employed in a different area (within a 3-year window). Column 6 shows that about half of the reduction in geographic inflows is due to a reduction of short-range movements, between municipalities within the same district. Estimates in columns 7 and 8 suggest further that the increase in outflows is mainly driven by increased outflows into non-employment. Outflows of native workers into employment in other areas are not found to increase in response to Czech inflows.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

This suggests that total employment responses to the immigration-induced supply shock are mainly driven by non-employed natives not entering employment in the affected area, although they would have done so in the absence of immigration. In addition, some workers who would have switched from other areas into the local labor market in the absence of immigration refrain from doing so. The share of native workers exiting into non-employment – the common interpretation of immigration-induced employment responses of native workers – is relatively small in magnitude. The results suggest further that spatial spillovers and in particular spillovers to neighboring areas (which in our setting would be particularly problematic), are modest, at least in the short run.

We have shown above that labor supply elasticities differ for workers in different skill- and age groups, and we note in Section 2 that the intensity of responses into and out of non-employment, and across areas will differ according to opportunity costs. These costs are in turn likely to differ across age- and skill groups, which may lead to different responses in inflows and outflows across the different groups. To investigate this further, we perform the de-composition in Panel A by skill (Panel B) and age groups (Panel C).

For unskilled workers, the overall employment response is large, as we discuss above, and nearly entirely driven by decreased inflows. Columns 3-6 show that this decrease is mainly driven by a reduction in inflows from non-employment. For skilled workers, the same pattern hold, but a reduction of inflows from other areas as well as increase of outflows into non-employment are likewise important, although smaller in magnitude. Geographic inflows are reduced mainly because fewer workers from far-away areas enter employment in exposed areas. In contrast, for unskilled workers, the reduction in geographic inflows is nearly entirely due to decreased inflows from nearby areas.

Interestingly, we find not only geographic inflows but also *outflows* of native workers to decrease in response to the entering of Czech workers. The likely reason is their particular distribution – areas along the border receive a particularly high share of Czech commuters, but as their surroundings are likewise in close distance to the border, they tend to also receive high shares. Geographic mobility over short range thus becomes unattractive for natives, in particular unskilled workers, whose movements are far more local than for skilled workers. This finding illustrates that the channels via which aggregate employment adjusts may depend

on the spatial distribution of the immigration-induced labor supply shock, and are therefore likely to differ across studies.

When we distinguish by different age classes, a distinct pattern evolves, where the employment effect of middle-aged workers is mainly driven by a reduction of inflows from non-employment, while outflows into non-employment are small and insignificant. On the other hand, while effects for older and younger workers are still mainly driven by a reduction in inflows, increase of outflows into non-employment are also important, in particular in the older age group. Thus, we find that the employment response to Czech inflows differs across age groups not only in magnitude, but also in the channels via which the aggregate level adjusts.

5.6 Discussion and Conclusions

A large literature examines the impact of immigration-induced labor supply shocks on native workers' labor market outcomes. However, there is little agreement on the degree to which these have adverse effects on employment and wages of natives. This lack of consensus is partly due to methodological difficulties that plague this literature – immigrants self-select into areas, responses of natives may be selective, and may dissipate their effect through the economy, and those responses are difficult to distinguish from other sources for area- or skill-specific trends.

In this paper we revisit this issue. We develop a methodological framework that allows responses of native workers to depend on group specific labor supply elasticities – an extension that we confirm to be important for the interpretation of estimates. We base our analysis on a uniquely suited natural experiment in Germany and the availability of detailed, population-wide panel data on individual workers, where in the aftermath of the fall of the iron curtain the German border region with the Czech Republic experienced an unexpected, sudden, and large inflow of Czech workers during the early 1990s. This supply shock had a distinctive and, for our purposes, useful feature – Czech workers were required to *commute* daily from their residence in the Czech Republic. This not only constitutes a unique experiment creating a *pure* labor supply shock, but it also allows us to use distance to border as additional exogenous variation for the magnitude of the labor supply shock affecting the eligible border areas.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

Studying the short-run response, we document a modest but rapid decline in wages, and a large decline in the employment of native workers. The observed response supports one of the core hypothesis held by the literature: as of imperfect substitutability, the negative effect of immigration falls most strongly on those workers whose qualifications become relatively more abundant. In our setting, the inflow of predominantly unskilled Czech workers led to particularly negative responses in employment and wages for unskilled native workers.

However, our findings are in stark contrast to other assumptions that are standard in the literature. First, we find both skilled and unskilled workers to be negatively affected – in contrast to the strand of literature that assume immigration to have (or that can identify) only distributionary, but no average effects. More importantly, we find that the notion of imperfect substitutability, while useful, does not suffice to explain the distributionary effect of immigration. In particular, we find the decline in employment to be strongest among older native workers, even though only few Czech commuters enter that age group.

Within our theoretical model, we can explain that finding by variation in workers' responsiveness to wage changes, that is, heterogeneity in the elasticity of labor supply across groups. This innovation, while consistent with evidence from other literatures, overturns standard assumptions on the distribution of the wage effect of immigration – with potentially severe implications, as these assumptions are directly used for the identification in one of the two dominant strands of the literature, the “skill cell” approach. Our results suggest that while differences in immigrant density across education groups can indeed be exploited, extension of that strategy along other dimensions, such as age or experience, are fragile, as workers are likely to respond differently to wage changes along those dimensions.

Overall, our results suggest that immigration had a large effect on native labor market outcomes. However, the results are less dramatic upon closer inspection. First, we focused on the short run, and the region under consideration is rural and received only few migrants before the fall of the iron curtain. Such areas may be less able to absorb large immigrant inflows than, for example, Miami or California. The particularities of the observed supply shock are also important. The inflow of Czech workers was unexpected, sudden, and of exceptional magnitude. Such shocks are likely to have far more negative effects than the gradual and potentially anticipated arrival of a smaller number of migrants. Second, Czech

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

workers had to commute, such that their contribution to local consumption was comparatively small. Indeed, we find that local population levels decreased in response to Czech inflows.

As such, the numbers presented in this study cannot be simply generalized to any type of immigration-induced supply shock. Instead, our estimates reflect the result from an atypical, but highly informative experiment – an experiment that in our view provides a more direct test of the core theoretical models in the literature than previous empirical work. Theoretical work almost always considers sudden, unexpected supply shocks that occur in isolation – ignoring the fact that immigrants may also shift local demand. Our “pure” supply shock comes closer to that stylized thought experiment than other natural experiments that have been studied in the literature. Our results suggest that such shock can indeed have substantially negative, both absolute and distributionary, effects on native workers. This result may come as a relief to those who argue that standard theories of factor demand are simply not consistent with studies that find immigration to have no negative effects on natives even in the short run.

Our findings may thus help to reconcile some of the existing arguments and strands in the literature. However, the availability of population-wide of our data allows us also to extend the literature in various new directions, and to study aspects of the adjustment process in local labor markets about which only little is known – the mechanisms via which the level of employment adjusts, the timing of those processes, and the degree to which they matter for different groups of workers.

First, we find the decline in native employment to be predominantly explained by movements from and to non-employment, while geographic movements across areas are far less important. The relative importance of the two channels varies with education and age, with young or skilled natives being more geographically mobile than older and unskilled workers.

Second, we decompose the overall employment effect further into changes in inflows – native workers who enter local employment – and changes in outflows of existing workers. Our findings suggest that inflows explain a far greater share of the total employment response than outflows, in stark contrast to the way employment responses to immigration are commonly interpreted. Inflows respond also more rapidly than outflows, which may explain why employment can adjust so rapidly – the adverse effect on native employment achieves full strength only one year after the inflow of Czech workers reached its peak. Finally, we find the

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

relative importance of the various inflow and outflow channels to differ in distinct and predictive ways with age and skill.

These findings are interesting not only for the immigration literature, but more generally, as they shed light on the channels via which labor markets, and native workers, adjust to local shocks. Our analysis provides a coherent picture of these processes, but more work is needed to assess their implications. For example, as each process affects different groups of workers, the welfare implications of immigration may depend crucially on what channel dominates in the adjustment of employment levels. Moreover, it seems likely that the importance of some processes depends strongly on the institutional environment, such as employment protection laws. The ways via which labor markets adjust to immigration, and the type of workers that are predominantly affected, might then differ across countries.

Appendix A: Derivation of Equilibrium Wage and Employment Responses

The equilibrium wage and employment response is determined by the two skill-specific labor demand curves,

$$d\log w_U = \varphi d\log L + (\beta - 1)(d\log L_U - d\log L) \quad (A1)$$

$$d\log w_S = \varphi d\log L + (\beta - 1)(d\log L_S - d\log L) \quad (A2)$$

where $d\log L = s_U \left(d\log L_U^N + dI \frac{\pi_U^I}{\pi_U^N} \right) + s_H \left(d\log L_S^N + dI \frac{\pi_S^I}{\pi_S^N} \right)$, and

$$d\log L_U^N = \eta_U d\log w_U \quad (A3)$$

$$d\log L_S^N = \eta_S d\log w_S. \quad (A4)$$

By plugging (A3) and (A4) into (A1) and (A2) we have

$$\begin{aligned} d\log w_U &= \varphi \left((s_U \eta_U d\log w_U + s_S \eta_S d\log w_S) + \Pi dI \right) \\ &\quad + (\beta - 1) \left(\eta_U d\log w_U - (s_U \eta_U d\log w_U + s_S \eta_S d\log w_S) + \left(\frac{\pi_U^I}{\pi_U^N} - \Pi \right) dI \right) \end{aligned} \quad (A5)$$

$$\begin{aligned} d\log w_S &= \varphi \left((s_U \eta_U d\log w_U + s_S \eta_S d\log w_S) + \Pi dI \right) \\ &\quad + (\beta - 1) \left(\eta_S d\log w_S - (s_U \eta_U d\log w_U + s_S \eta_S d\log w_S) + \left(\frac{\pi_S^I}{\pi_S^N} - \Pi \right) dI \right) \end{aligned} \quad (A6)$$

Solving (A5) and (A6) for $d\log w_U$ and $d\log w_S$, respectively, we have

$$\begin{aligned} &d\log w_U \\ &= \frac{(\varphi - (\beta - 1))s_S \eta_S d\log w_S + \varphi \Pi dI + (\beta - 1) \left(\frac{\pi_U^I}{\pi_U^N} - \Pi \right) dI}{1 - \varphi s_U \eta_U - (\beta - 1)s_S \eta_U} \end{aligned} \quad (A7)$$

$$\begin{aligned} &d\log w_S \\ &= \frac{(\varphi - (\beta - 1))s_U \eta_U d\log w_U + \varphi \Pi dI + (\beta - 1) \left(\frac{\pi_S^I}{\pi_S^N} - \Pi \right) dI}{1 - \varphi s_S \eta_S - (\beta - 1)s_U \eta_S}. \end{aligned} \quad (A8)$$

Plugging (A8) into (A7) and bringing all terms on a common denominator we have

$$\begin{aligned} &d\log w_U \\ &= \frac{(\varphi - (\beta - 1))^2 s_U s_S \eta_U \eta_S d\log w_U + (\varphi - (\beta - 1))s_S \eta_S \left(\varphi \Pi + (\beta - 1) \left(\frac{\pi_U^I}{\pi_U^N} - \Pi \right) \right)}{(1 - \varphi s_U \eta_U - (\beta - 1)s_S \eta_U)(1 - \varphi s_S \eta_S - (\beta - 1)s_U \eta_S)} dI \end{aligned}$$

$$+ \frac{(1 - \varphi s_S \sigma_S - (\beta - 1)s_U \eta_S)\varphi \Pi + (1 - \varphi s_S \eta_S - (\beta - 1)s_U \eta_S)(\beta - 1)\left(\frac{\pi_U^I}{\pi_U^N} - \Pi\right)}{(1 - \varphi s_U \eta_U - (\beta - 1)s_S \eta_U)(1 - \varphi s_S \eta_S - (\beta - 1)s_U \eta_S)} dI.$$

Solving for $d\log w_U$ yields

$$\begin{aligned} d\log w_L &= \frac{(\varphi - (\beta - 1))s_S \eta_S \left(\varphi \Pi + (\beta - 1)\left(\frac{\pi_S^I}{\pi_S^N} - \Pi\right) \right)}{(1 - \varphi s_U \eta_U - (\beta - 1)s_S \eta_U)(1 - \varphi s_S \eta_S - (\beta - 1)s_U \eta_S) - (\varphi - (\beta - 1))^2 s_U s_S \eta_U \eta_S} dI \\ &+ \frac{(1 - \varphi s_S \eta_S - (\beta - 1)s_U \eta_S)\varphi \Pi + (1 - \varphi s_S \eta_S - (\beta - 1)s_U \eta_S)(\beta - 1)\left(\frac{\pi_U^I}{\pi_U^N} - \Pi\right)}{(1 - \varphi s_U \eta_U - (\beta - 1)s_S \eta_U)(1 - \varphi s_S \eta_S - (\beta - 1)s_U \eta_S) - (\varphi - (\beta - 1))^2 s_U s_S \eta_U \eta_S} dI \end{aligned}$$

Simplifying both the numerators and the denominator, and using $\phi = \frac{\varphi}{\beta - 1} - 1$, we find that the wage response equals

$$\begin{aligned} d\log w_U &= \frac{(\beta - 1) \left[\frac{\pi_U^I}{\pi_U^N} (1 - \varphi \eta_S) - \Pi \left(1 - \frac{\varphi}{\beta - 1}\right) \right]}{1 - (\beta - 1)[\eta_U(1 + s_U \phi) + \eta_S(1 + s_S \phi) - \eta_U \eta_S(1 + \phi)(\beta - 1)]} dI \end{aligned} \quad (A9)$$

Appendix B: Wage and Employment Responses with Three Skill

Groups

We now extend the model to three skill groups, but impose the restriction that capital is fully elastic (i.e., $\varphi = 0$). With three skill groups, labor L is a CES aggregate of low (L), medium (M), and high (H) skilled labor, such that

$$L = \left[\theta_L L_L^\beta + \theta_M L_M^\beta + \theta_H L_H^\beta \right]^{\frac{1}{\beta}} \quad (B1)$$

As before, we have $d\log L_g = \frac{\pi_g^I}{\pi_g^N} dI + \eta_g d\log w_g$, see equation (4b), while equation (5) becomes

$$d\log L = \Pi dI + s_L \eta_L d\log w_L + s_M \eta_M d\log w_M + s_H \eta_H d\log w_H,$$

with $s_g = \frac{\theta_g L_g^\beta}{\theta_L L_L^\beta + \theta_M L_M^\beta + \theta_H L_H^\beta}$. Since $\varphi = 0$, totally differentiating equation (8) yields

$$d\log w_g = (\beta - 1)(d\log L_g - d\log L). \quad (B2)$$

Plugging in the expressions for $d\log L_g$ and $d\log L$, and solving for $d\log w_g$ we obtain, for $g = L$,

$$d\log w_L = \frac{(\beta - 1) \left((s_M(1 - (\beta - 1)\eta_H) \left(\frac{\pi_L^I}{\pi_L^N} - \frac{\pi_M^I}{\pi_M^N} \right) + s_H(1 - (\beta - 1)\eta_M) \left(\frac{\pi_L^I}{\pi_L^N} - \frac{\pi_H^I}{\pi_H^N} \right) \right)}{1 - (\beta - 1)\#1 + (\beta - 1)^2\#2} dI$$

where

$$\#1 = ((1 - s_L)\eta_L + (1 - s_M)\eta_M + (1 - s_H)\eta_H)$$

$$\#2 = ((1 - s_L - s_M)\eta_L\eta_M + (1 - s_M - s_H)\eta_M\eta_H + (1 - s_L - s_H)\eta_L\eta_H).$$

The employment response follows from

$$d\log L_L^N = \eta_L d\log w_L$$

“Perverse” wage effects are possible. Suppose that

$$\frac{\pi_L^I}{\pi_L^N} > \frac{\pi_M^I}{\pi_M^N} > \frac{\pi_H^I}{\pi_H^N}$$

that is, migrant concentration high in skill group L , medium in skill group M , and low in skill group H . It is nevertheless possible that wages of the medium skilled decline more than wages of the unskilled (i.e., $d\log w_M < d\log w_L$) if the local labor supply elasticity of the medium skilled is large relative to that of the unskilled. It is, however, not possible that wages of the skilled (which must increase if capital is fully flexible) decline relative to wages of the unskilled (which decline).

Appendix C: The Commuting Policy

Various schemes for legal employment of foreign nationals in Germany were extended or introduced with effect to January 1st, 1991.³⁶ The provision introduced new nationwide immigration rules such as the controversial “*Werkvertragsregelung*”, which granted firms the right to contract foreign workers for specific work assignments. But it also comprised a locally constrained scheme that received little public attention: the commuting scheme “*Grenzgängerregelung*”, which granted foreign nationals from neighboring countries the right to work in dependent employment in German border regions. However, it did not grant residency; *Grenzgänger* were required to commute daily from their country of origin *or* to work for a maximum of two days per week. The policy was otherwise non-restrictive, e.g. it was not constrained to specific industries or applicants with specific qualifications. Since the movement of labor was in principal unrestricted within the European Economic Community (bilateral agreements already covered tax and other issues on the western borders), this policy

³⁶ See “Anwerbestoppausnahme-Verordnung”, Bundesgesetzblatt, Jahrgang 1990, Teil I.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

had consequences only at the eastern German borders to Poland and Czechoslovakia (from 1993 Czech Republic).³⁷ We restrict our analysis to the West-German border region with the Czech Republic because social security records for the workforce in East Germany are only fully available from 1992 onwards, and because the East was subject to strong institutional and economic transitions after reunification. The intended implementation of the policy along the Czech-German border was first reported in September 1990, only shortly before the scheme came into effect.³⁸

Figure 1 maps the region that was affected by the scheme, comprising 21 districts that lie within an approximate eighty kilometers band from the Czech-German border. The initial provision lists only 18 districts explicitly, and does not include the districts *Straubing*, *Deggendorf*, and *Straubing-Bogen*. However, these districts were in a similar distance from the border as other, listed, districts, and experienced a similar inflow of Czech workers. Moreover, the districts are listed in the revised provision from 1998. We thus consider all 21 districts as treated. Our results are robust to the exclusion of the three districts in question.

After German reunification in 1990, districts close to the former inner-German border were also affected by commuters from East Germany. To avoid contamination we thus exclude all districts located within about eighty kilometres of the inner-German border. This choice is motivated by external data on regional commuting flows from the late 1990s, which shows that areas directly adjacent to the inner German border received a high share of commuters;³⁹ and, from our own data, by the spatial distribution of newly registered workers, which in 1990 and 1991 are substantial higher as East-Germans entered West-German labor markets for the first time. Thirteen from 21 districts remain, as illustrated in Figure 1. We use the term *border region* to refer to the remaining districts. The policy remained in effect throughout the time

³⁷ A summary of the existing commuting schemes within the European Community is given in IAB (1993), *Mitteilungen aus der Arbeitsmarkt und Berufsforschung* 26/93, “Beschäftigung von Grenzarbeitnehmern in der Bundesrepublik Deutschland”.

³⁸ See *Süddeutsche Zeitung*, 1./2.9.1990. Implementation of the scheme along the Czech-German border occurred shortly before its general introduction on January 1, 1991 (see Hönekopp, “Ost-West-Wanderungen: Die neuen Migrationsbewegungen”). A detailed examination of daily employment records confirms that the inflow of Czech nationals started already in September 1990.

³⁹ See for example Kropp (2010), “Veränderungen der Pendlerverflechtungen in Deutschland zwischen 1993 und 2008”. In: Troeger-Weiß and Jurczek (Eds.), “Nationale und regionale Mobilität von Unternehmen und Arbeitnehmern”, and IAB (1992), *Mitteilungen aus der Arbeitsmarkt und Berufsforschung* 25/92, “Ostdeutsche Arbeitskräfte in Westdeutschland”.

CHAPTER 5. THE IMPACT OF IMMIGRATION ON LOCAL LABOR MARKETS

period studied in this paper, with minor changes in subsequent revisions of the legislation.⁴⁰ The substantial increase of foreign workers in Germany in general and of the Czech commuters in the border region was perceived to have negative consequences for native workers. Commentators deplored in particular that Czech and other foreign workers were employed at wages far below the prevailing wage level for German workers, and that worker protection laws and rights were circumvented. However, the main political backlash was directed not against the locally constraint commuting schemes, but against the temporary employment of foreign workers in German firms under the national “*Werkvertragsregelung*” scheme.⁴¹

⁴⁰ See “*Anwerbestoppausnahme-Verordnung*”, Bundesgesetzblatt, Jahrgang 1998, Teil I, §6. The revision restricted the eligibility of marginally employed and individuals who receive social benefits in their country of origin.

⁴¹ See for example the motion against its abuse by the Social Democratic Party in the German parliament, Deutscher Bundestag, Drucksache 12/3299, 23.9.1992.

Tables and Figures

Table 5.1: Characteristics of Treated, Inland and Matched Control Districts

		border	all inland	matched inland
<i>skill</i>	low	0.274	0.229	0.244
	medium	0.695	0.703	0.723
	high	0.030	0.069	0.034
<i>age</i>	below 30	0.434	0.351	0.420
	30 to 49	0.410	0.454	0.412
	50 and above	0.157	0.195	0.168
	female	0.411	0.401	0.414
	foreign	0.025	0.081	0.035
	mean log wages (censored)	3.881	4.055	3.879
	share censored	0.023	0.048	0.027
	number of districts	13	329	24

Note: The table compares average characteristics (weighted by employment level) in 1989 in eligible districts in the border region; all other West-German districts; and matched inland districts (see Figure 1). Low-skilled workers have no postsecondary education. Medium-skilled workers completed an apprenticeship or a high school degree (Abitur). High-skilled workers graduated from a university or college (Universität or Fachhochschule). Longitudinal information used to impute education variable, following Fitzenberger et al. (2006). Remaining missings (3.9 percent) classified as low skilled.

Table 5.2: Characteristics of Czech and Non-Czech Nationals in Border Region

		Non-Czech	Czech
<i>skill distribution</i>			
	low skilled	0.276	0.505
	high skilled	0.724	0.495
<i>age distribution</i>			
	below 30	0.435	0.344
	30-49	0.408	0.619
	50 and above	0.157	0.0368
<i>age distribution: low skilled</i>			
	below 30	0.138	0.187
	30-49	0.08	0.2995
	50 and above	0.0577	0.0185
<i>age distribution: high skilled</i>			
	below 30	0.297	0.157
	30-49	0.328	0.32
	50 and above	0.0993	0.0183
	share female	0.411	0.163
	mean log wages (imputed)	4.085	3.850
<i>industries</i>			
	public sector	0.171	0.0213
	pit and quarry	0.0273	0.0479
	wood processing	0.0323	0.0736
	construction	0.0987	0.249
	hotels and restaurants	0.0299	0.0924
	N	323,039	9,996

Note: The table compares average characteristics of Czech commuters (in 1992) against the pre-existing, non-Czech workforce (in 1989).

Table 5.3: First Stage

	$\frac{Czech_{1992} - Czech_{1990}}{Employment_{1990}}$
distance	-0.00338 (.000860)
distance squared	0.0000268 (.0000103)
constant	0.115 (.0160)
N	290
R-sq	0.387
F	48.97

Note: The table reports coefficient estimates from the first stage regression of the growth in the employment share of Czech workers on airline distance (km) and distance squared to the next border crossing. Estimated across municipalities within the border region, weighted by employment in 1990. Pooling also over inland areas, and interacting the distance variables with an indicator variable equal to 1 if a municipality is part of the border region, yields an R-squared of 0.544 and a F-statistic of 57.07.

Table 5.4: Wage and Employment Baseline Estimates by Skill

		employment	wages	local labor supply elasticity
<i>Panel A: all</i>				
(i)	coef 91 on $\frac{Czech_{1991} - Czech_{1990}}{Employment_{1990}}$	-0.425 (0.265)	-0.160 (0.053)	2.661
(ii)	coef 91 + 92 on $\frac{Czech_{1992} - Czech_{1990}}{Employment_{1990}}$	-0.620 (0.198)	-0.121 (0.037)	5.130
(iii)	coef 91 + 92 + 93 on $\frac{Czech_{1992} - Czech_{1990}}{Employment_{1990}}$	-0.926 (0.247)	-0.134 (0.051)	6.885
(iv)	as (iii), but OLS	-0.289 (0.143)	-0.059 (0.036)	
<i>Panel B: low-skilled</i>				
(i)	coef 91 on $\frac{Czech_{1991} - Czech_{1990}}{Employment_{1990}}$	-1.087 (0.402)	-0.176 (0.080)	6.177
(ii)	coef 91 + 92 on $\frac{Czech_{1992} - Czech_{1990}}{Employment_{1990}}$	-1.022 (0.257)	-0.169 (0.053)	6.034
(iii)	coef 91 + 92 + 93 on $\frac{Czech_{1992} - Czech_{1990}}{Employment_{1990}}$	-1.371 (0.322)	-0.202 (0.064)	6.794
(iv)	as (iii), but OLS	-0.805 (0.182)	-0.095 (0.051)	
<i>Panel C: high-skilled</i>				
(i)	coef 91 on $\frac{Czech_{1991} - Czech_{1990}}{Employment_{1990}}$	-0.058 (0.246)	-0.148 (0.052)	0.394
(ii)	coef 91 + 92 on $\frac{Czech_{1992} - Czech_{1990}}{Employment_{1990}}$	-0.313 (0.182)	-0.102 (0.038)	3.062
(iii)	coef 91 + 92 + 93 on $\frac{Czech_{1992} - Czech_{1990}}{Employment_{1990}}$	-0.501 (0.224)	-0.106 (0.053)	4.712
(iv)	as (iii), but OLS	0.020 (0.142)	-0.055 (0.037)	

Note: The Table is based on coefficient estimates from the 2SLS (rows i-iii) or OLS (row iv) regression of yearly employment or wage growth of natives (Panel A) or natives in a specific skill group (Panel B and C) on the growth in the employment share of Czech workers in the municipality, see equations (17) and (18). Yearly coefficient estimates are added to show the cumulative effect, relative to the 1990 baseline. See Table 3 for first stage. Each observation is weighted by the group-specific number of native workers employed in the base year in that particular municipality. Standard errors are computed using the bootstrap method (500 replications, resampling on the municipality level).

Table 5.5: Robustness Checks

<i>Panel A: Wage Effects</i>		(i)	(ii)	(iii)	(iv)	(v)
		Baseline	trend-adjusted	long differences	inner border vs. inland	border only
	all	-0.131 (0.049)	-0.178 (0.034)	0.002 (0.052)	-0.134 (0.047)	-0.124 (0.103)
	low skilled	-0.178 (0.064)	-0.224 (0.064)	-0.057 (0.088)	-0.189 (0.060)	-0.250 (0.124)
	high skilled	-0.115 (0.052)	-0.170 (0.034)	-0.052 (0.052)	-0.115 (0.050)	-0.100 (0.103)
<i>Panel B: Employment Effects</i>		(i)	(ii)	(iii)	(iv)	(v)
		Baseline	trend-adjusted	long differences	inner border vs. inland	border only
	all	-0.926 (0.224)	-0.927 (0.287)	-0.930 (0.260)	-0.963 (0.241)	-0.952 (0.338)
	low skilled	-1.371 (0.308)	-1.417 (0.417)	-1.218 (0.288)	-1.411 (0.304)	-1.036 (0.410)
	high skilled	-0.501 (0.216)	-0.866 (0.280)	-0.521 (0.245)	-0.550 (0.230)	-0.586 (0.362)

Note: The table presents coefficient estimates from various robustness tests. Column 1 reports our baseline estimates (see Table 4). Column 2 reports estimates from pooled (over years 1987-1993) regressions, in which the pre-treatment observations identify municipality-specific differences in trend. In column 3 we report estimates for which we take long differences (between 1990 and 1993) instead of accumulating yearly coefficients (employment and wage coefficients) and average wages over all observed workers (wage coefficients). For column 5 we pool municipalities from the inland and that half of the border region that is closer to the Czech-German border, and which received the majority of Czech commuters. For column 6 we pool municipalities within the border region only.

Table 5.6: Employment and Wage Effects by Skill and Age Groups

		<u>unadjusted</u>			<u>trend-adjusted</u>		
		employment	wages	local labor supply elasticity	employment	wages	local labor supply elasticity
<i>Panel A: all</i>							
	below 30	-0.832	-0.316	2.64	-0.604	-0.305	1.98
	(share Czechs: 0.031)	(0.325)	(0.079)		(0.392)	(0.066)	
	30 to 49	-0.534	-0.100	5.36	-0.964	-0.147	6.54
	(share Czechs: 0.040)	(0.259)	(0.055)		(0.321)	(0.069)	
	50 and above	-1.945	-0.068	28.65	-1.428	-0.172	8.32
	(share Czechs: 0.007)	(0.359)	(0.048)		(0.417)	(0.057)	
<i>test LS elasticity (one-sided p-value):</i>							
(1) 30 to 49 >= below 30				0.888			0.974
(2) 50 and above >= below 30				0.904			0.994
<i>Panel B: low-skilled</i>							
	below 30	-2.262	-0.558	4.05	-1.601	-0.441	3.63
	(share Czechs: 0.112)	(0.573)	(0.104)		(0.663)	(0.130)	
	30 to 49	-0.704	-0.179	3.94	-1.428	-0.237	6.02
	(share Czechs: 0.107)	(0.349)	(0.087)		(0.476)	(0.121)	
	50 and above	-1.364	-0.097	14.10	-1.324	-0.194	6.82
	(share Czechs: 0.011)	(0.388)	(0.073)		(0.502)	(0.088)	
<i>test LS elasticity (one-sided p-value):</i>							
(1) 30 to 49 >= below 30				0.470			0.786
(2) 50 and above >= below 30				0.904			0.836
<i>Panel C: high-skilled</i>							
	below 30	-0.283	-0.276	1.02	-0.457	-0.281	1.63
	(share Czechs: 0.017)	(0.295)	(0.085)		(0.394)	(0.068)	
	30 to 49	-0.191	-0.090	2.12	-1.012	-0.142	7.13
	(share Czechs: 0.025)	(0.252)	(0.057)		(0.332)	(0.061)	
	50 and above	-1.636	-0.066	24.72	-1.337	-0.158	8.47
	(share Czechs: 0.005)	(0.336)	(0.050)		(0.434)	(0.058)	
<i>test LS elasticity (one-sided p-value):</i>							
(1) 30 to 49 >= below 30				0.674			0.988
(2) 50 and above >= below 30				0.890			0.994

Note: The Table is based on coefficient estimates from the 2SLS regression of yearly employment or wage growth of natives (Panel A) or natives in a specific skill group (Panel B and C) on the growth in the employment share of Czech workers in the municipality, see equations (17) and (18). Yearly coefficient estimates for 1991, 1992 and 1993 are added to show the cumulative effect relative to 1990. Trend-adjusted estimates are from a pooled regression over years 1987-1993, in which the pre-treatment observations are used to identify municipality-specific differences in trend. See Table 3 for first stage. Each observation is weighted by the group-specific number of native workers employed in the base year in that particular municipality. Standard errors are computed using the bootstrap method (500 replications, resampling on the municipality level). Test statistics (1) and (2) are computed as the proportion of bootstrap samples (500 repetitions) where the LS elasticity of natives aged 30 to 49 (aged 50 and above) was greater than or equal to the labor supply elasticity of natives below age 30.

Table 5.7: Margins of Adjustment: Unemployment versus Movements between Areas

	geographical vs non-employment			more detailed movements		
	total employment	geographical	non-employment	geographical, incl. through non-employment	geographical far, incl. employment	population
<i>Panel A: all</i>						
<i>share of baseline employment</i>	-0.989 (0.342)	0.06 -0.167 (0.229)	0.11 -0.821 (0.190)	0.08 -0.287 (0.256)	0.05 -0.158 (0.151)	-0.298 (0.065)
<i>Panel B: by skill</i>						
low skilled						
<i>share of baseline employment</i>	-1.256 (0.444)	0.04 0.033 (0.272)	0.12 -1.289 (0.298)	0.05 -0.049 (0.311)	0.03 0.033 (0.206)	
high skilled						
<i>share of baseline employment</i>	-0.875 (0.347)	0.06 -0.218 (0.238)	0.11 -0.658 (0.192)	0.08 -0.326 (0.263)	0.05 -0.191 (0.151)	
<i>Panel C: by age</i>						
below 30						
<i>share of baseline employment</i>	-0.555 (0.450)	0.09 -0.147 (0.269)	0.17 -0.560 (0.274)	0.12 -0.254 (0.322)	0.07 -0.137 (0.182)	
30 to 49						
<i>share of baseline employment</i>	-1.180 (0.371)	0.05 -0.257 (0.259)	0.07 -0.857 (0.183)	0.06 -0.381 (0.279)	0.04 -0.231 (0.179)	
50 and above						
<i>share of baseline employment</i>	-1.349 (0.416)	0.02 -0.050 (0.201)	0.10 -1.203 (0.275)	0.03 -0.061 (0.228)	0.02 0.043 (0.133)	

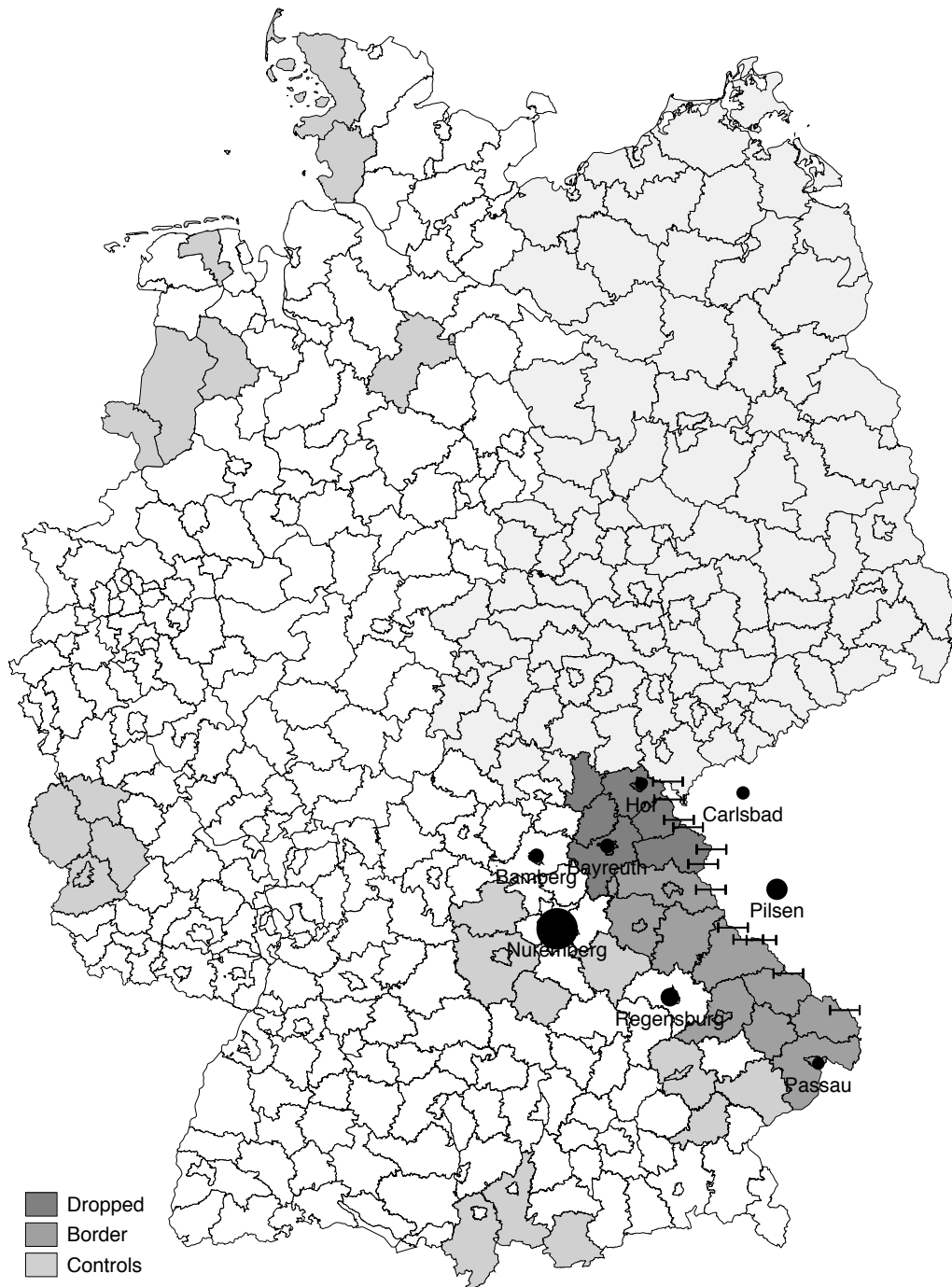
Note: Coefficients reported in the table are equal to the inflow minus the outflow coefficient in the respective employment category (see Table 8). See Table 3 for first stage and main text for a description of each category. Estimation is pooled over years 1989-1993 and trend-corrected. The yearly coefficient estimates for 1991, 1992 and 1993 are added to show the cumulative effect relative to 1990, and pre-treatment observations identify municipality-specific differences in trend. Standard errors are computed using the bootstrap method (500 replications, resampling on the municipality level). The share of baseline employment are equal to the average of the respective inflow and outflows shares in 1985-1989; for example, the yearly average share of new entrants arriving from employment in other municipalities is 5.8 percent, the yearly share of workers leaving into other areas is 5.9 percent, and the average is 5.85 percent.

Table 5.8: Inflows vs. Outflows

	<u>total</u>		<u>by type: inflows</u>				<u>by type: outflows</u>			
	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	(x)
	inflows	outflows	non-employment	geographical	incl. through non-emp.	geographical close, direct	non-employment	geographical	incl. through non-emp.	geographical close, direct
<i>Panel A: all</i>	-0.878 (0.258)	0.111 (0.161)	-0.645 (0.178)	-0.233 (0.172)	-0.268 (0.185)	-0.130 (0.120)	0.176 (0.084)	-0.066 (0.154)	0.019 (0.168)	-0.063 (0.117)
<i>Panel B: by skill</i>										
low skilled	-1.385 (0.341)	-0.129 (0.238)	-1.154 (0.261)	-0.231 (0.218)	-0.289 (0.235)	-0.221 (0.130)	0.135 (0.151)	-0.264 (0.180)	-0.240 (0.193)	-0.161 (0.131)
high skilled	-0.761 (0.259)	0.115 (0.160)	-0.498 (0.179)	-0.263 (0.179)	-0.292 (0.191)	-0.121 (0.132)	0.159 (0.078)	-0.045 (0.160)	0.034 (0.176)	-0.043 (0.120)
<i>Panel C: by age</i>										
below 30	-0.594 (0.327)	-0.039 (0.201)	-0.315 (0.272)	-0.280 (0.188)	-0.273 (0.209)	-0.175 (0.131)	0.245 (0.111)	-0.133 (0.190)	-0.019 (0.220)	-0.096 (0.147)
30 to 49	-1.330 (0.297)	-0.150 (0.180)	-0.807 (0.171)	-0.275 (0.200)	-0.331 (0.208)	-0.114 (0.126)	0.050 (0.072)	-0.018 (0.167)	0.050 (0.172)	-0.040 (0.115)
50 and above	-0.974 (0.287)	0.375 (0.238)	-0.726 (0.163)	-0.051 (0.168)	-0.124 (0.188)	-0.067 (0.131)	0.477 (0.234)	-0.101 (0.134)	-0.063 (0.151)	-0.045 (0.107)

Note: The Table is based on coefficient estimates from the 2SLS regression of the yearly share of each employment category of natives (Panel A) or natives in a specific skill group (Panel B and C) on the growth in the employment share of Czech workers in the municipality. See Table 3 for first stage and main text for a description of each category. Estimation is pooled over years 1989-1993 and trend-corrected. The yearly coefficient estimates for 1991, 1992 and 1993 are added to show the cumulative effect relative to 1990, and pre-treatment observations identify municipality-specific differences in trend. Each observation is weighted by the group-specific number of native workers employed in the base year in that particular municipality. Standard errors are computed using the bootstrap method (500 replications, resampling on the municipality level).

Figure 5.1: Border Region



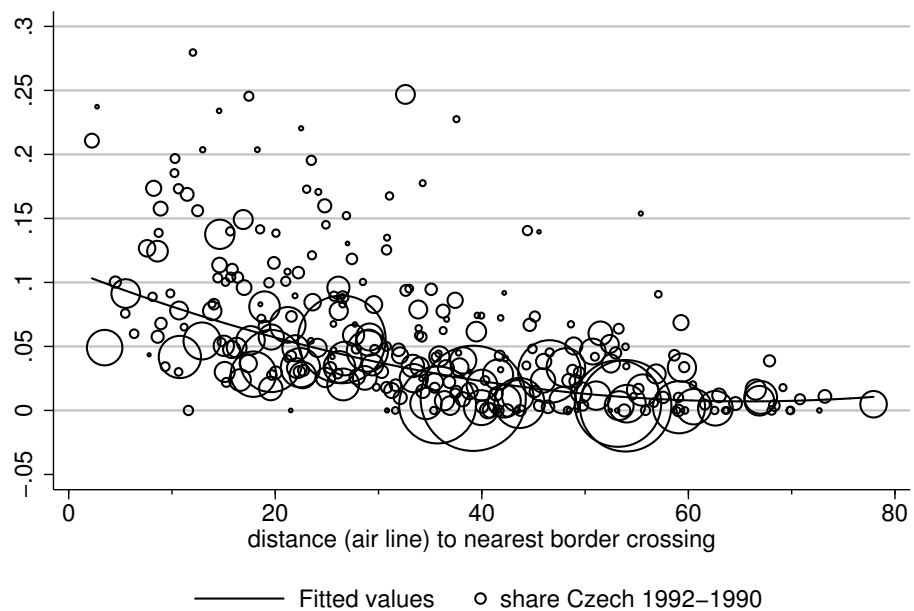
Note: The map depicts districts eligible under the commuting policy (dark and medium-dark grey), matched control districts (medium grey), and other districts in West and former East Germany (light grey). Eligible districts close to the inner German border (dark grey) are dropped in the analysis. The map further depicts crossings along and cities near the Czech-German border. Plot produced with the package "spmap" for Stata.

Figure 5.2: Employment Shares of Czech nationals: Border vs Inland



Note: The Figure plots the share of Czech workers in local employment in border region and matched inland districts (see Figure 1) before and after introduction of the commuting policy in 1991.

Figure 5.3: Spatial Distribution of Czech Nationals in Border Region



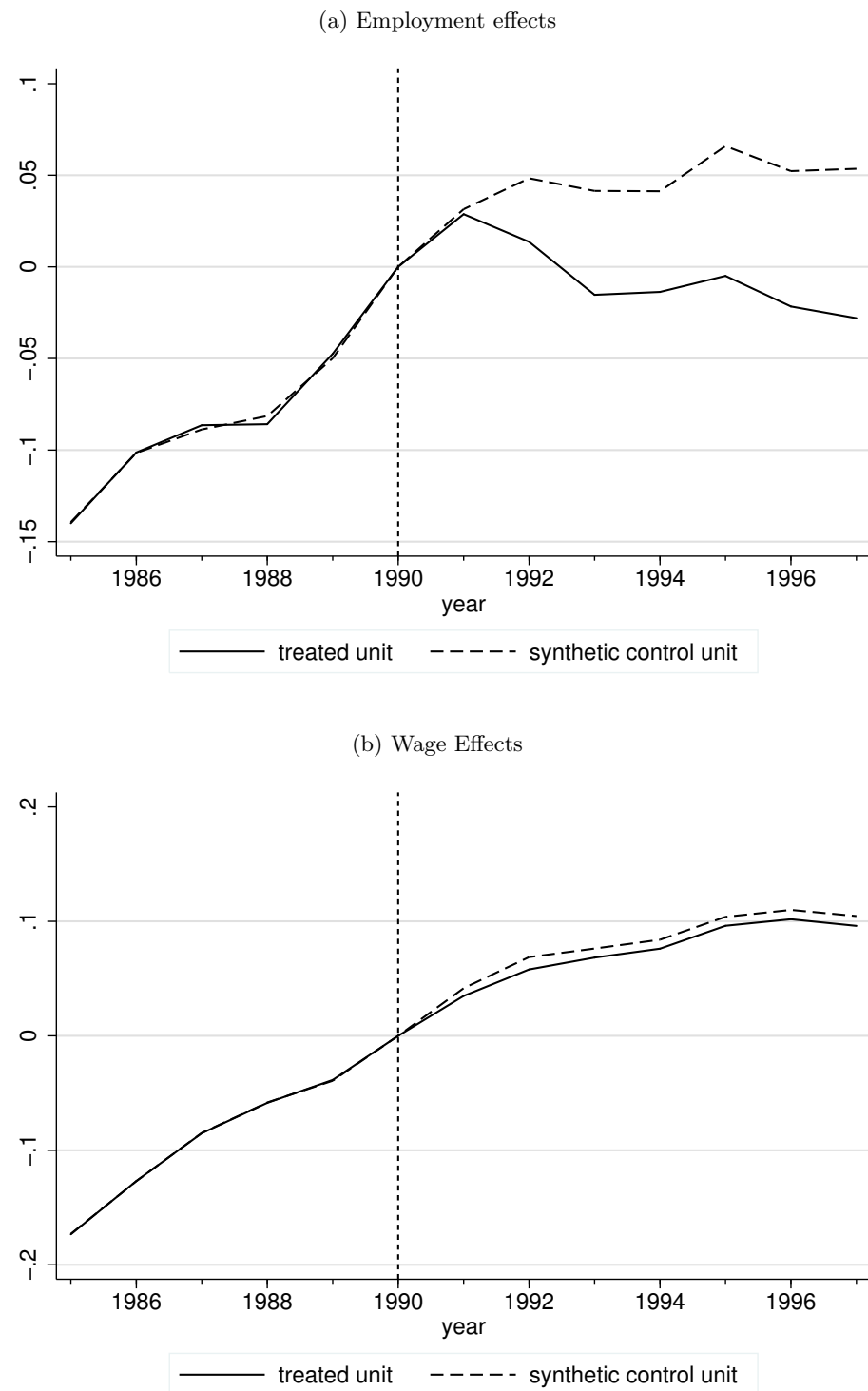
Note: The Figure plots, for each municipality within the eligible border region (see Figure 1), the increase in the number of Czech workers as a share of employment in 1990 against the airline distance of the centroid of the municipality to the closest border crossing. The size of each circle is proportional to employment in 1990. Fitted values are from a regression on distance and distance squared.

Figure 5.4: Aggregate Wage and Employment Effects



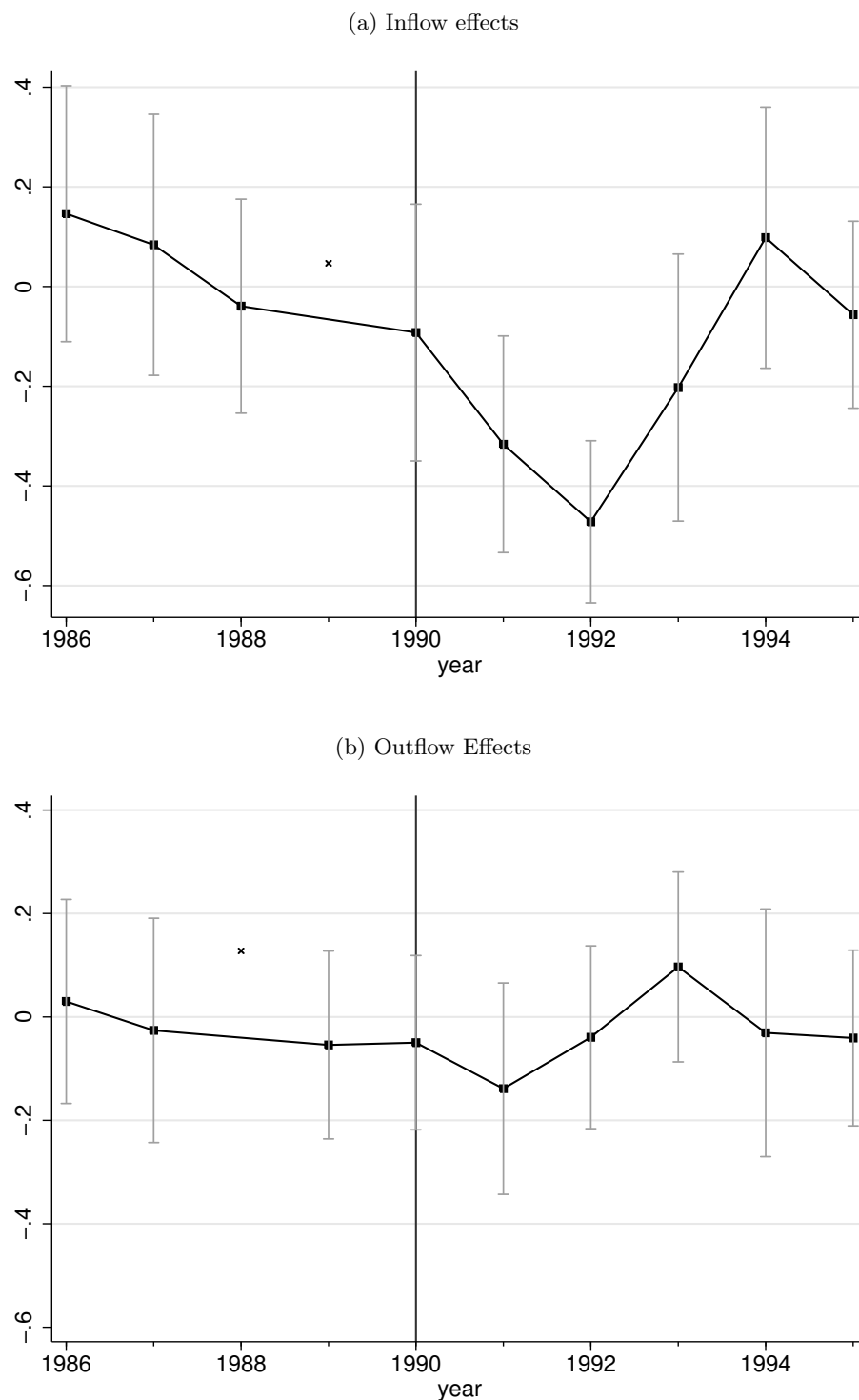
Note: The Figures are based on coefficient estimates from the 2SLS regressions of yearly employment or wage growth of natives on the growth in the employment share of Czech workers in the municipality, see equations (17) and (18). Yearly coefficient estimates are added to show the cumulative effect, relative to the 1990 baseline. See Table 3 for first stage. Each observation is weighted by the number of native workers employed in the base year in that particular municipality. The 95% confidence intervals are computed using the bootstrap method (500 replications, resampling on the municipality level).

Figure 5.5: Synthetic Control Method, Wage and Employment Effects



Note: Trends in wage and employment growth of native workers (relative to 1990 baseline), inner border region vs. synthetic control.

Figure 5.6: Yearly Native Inflow and Outflow Effects



Note: The Figures present coefficient estimates from the 2SLS regressions of yearly inflow or outflow rates of natives on the growth in the employment share of Czech workers in the municipality, see main text. See Table 3 for first stage. Each observation is weighted by the number of native workers employed in the base year in that particular municipality. The 95% confidence intervals are computed using the bootstrap method (500 replications, resampling on the municipality level). The coefficient estimate for outflows in 1989 and inflows in 1988 represent outliers (see main text) and are plotted, but not connected to other coefficient estimates.

Bibliography

- AARONSON, D., AND B. MAZUMDER (2008): “Intergenerational Economic Mobility in the United States, 1940 to 2000,” *Journal of Human Resources*, 43(1).
- ABADIE, A., A. DIAMOND, AND J. HAINMUELLER (2010): “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of Californias Tobacco Control Program,” *Journal of the American Statistical Association*, 105(490), 493–505.
- ABADIE, A., AND J. GARDEAZABAL (2003): “The Economic Costs of Conflict: A Case Study of the Basque Country,” *American Economic Review*, 93(1), 113–132.
- ADERMON, A. (2013): “Educational Mobility Across Three Generations,” Ph.D. thesis, Uppsala University.
- ANGER, S. (2011): “The Intergenerational Transmission of Cognitive and Non-Cognitive Skills During Adolescence and Young Adulthood,” IZA Discussion Papers 5749, Institute for the Study of Labor (IZA).
- ANGRIST, J., AND A. KRUEGER (1999): “Empirical Strategies in Labor Economics,” in *Handbook of Labor Economics 3A*, ed. by O. Ashenfelter, and D. Card, vol. 3A of *Handbook of Labor Economics*, chap. 23. Elsevier.
- ATKINSON, A. (1980): “On Intergenerational Income Mobility in Britain,” *Journal of Post Keynesian Economics*, 3(2), 194–218.
- ATKINSON, A. B., AND S. JENKINS (1984): “The Steady-State Assumption and the Estimation of Distributional and Related Models,” *Journal of Human Resources*, 19(3), 358–376.
- ATKINSON, A. B., T. PIKETTY, AND E. SAEZ (2011): “Top Incomes in the Long Run of History,” *Journal of Economic Literature*, 49(1), 3–71.

- AUTOR, D., AND L. KATZ (1999): “Changes in the Wage Structure and Earnings Inequality,” in *Handbook of Labor Economics*, ed. by O. C. Ashenfelter, and D. Card, vol. 3, pp. 1463–1555. North-Holland.
- AYDEMIR, A., AND G. J. BORJAS (2011): “Attenuation Bias in Measuring the Wage Impact of Immigration,” *Journal of Labor Economics*, 29(1), 69–113.
- BECKER, G., AND N. TOMES (1979): “An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility,” *The Journal of Political Economy*, 87(6), 1153–1189.
- (1986): “Human Capital and the Rise and Fall of Families,” *Journal of Labor Economics*, 4(3), 1–39.
- BERNARD, A. B., AND S. N. DURLAUF (1996): “Interpreting Tests of the Convergence Hypothesis,” *Journal of Econometrics*, 71(1-2), 161–173.
- BHULLER, M., M. MOGSTAD, AND K. G. SALVANES (2011): “Life-Cycle Bias and the Returns to Schooling in Current and Lifetime Earnings,” IZA Discussion Papers 5788, Institute for the Study of Labor (IZA).
- BJÖRKLUND, A. (1993): “A Comparison between Actual Distributions of Annual and Lifetime Income: Sweden 1951-89,” *Review of Income and Wealth*, 39(4), 377–386.
- BJÖRKLUND, A., AND M. JÄNTTI (2009): “Intergenerational Income Mobility and the Role of Family Background,” in *Oxford Handbook of Economic Inequality*, ed. by B. N. Salverda Wiemer, and T. M. Smeeding. Oxford University Press.
- BJÖRKLUND, A., M. JÄNTTI, AND M. J. LINDQUIST (2009): “Family Background and Income During the Rise of the Welfare State: Brother Correlations in Income for Swedish Men born 1932-1968,” *Journal of Public Economics*, 93(5-6), 671 – 680.
- BJÖRKLUND, A., M. LINDAHL, AND E. PLUG (2006): “The Origins of Intergenerational Associations: Lessons from Swedish Adoption Data,” *Quarterly Journal of Economics*, 121(3), 999–1029.
- BJÖRKLUND, A., AND K. SALVANES (2010): “Education and Family Background: Mechanisms and Policies,” in *Handbook of the Economics of Education*, ed. by S. M. E.A. Hanushek, and L. Woessmann, vol. 3.

- BLACK, S. E., AND P. DEVEREUX (2011): “Recent Developments in Intergenerational Mobility,” in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 4A. Elsevier.
- BLANDEN, J. (2009): “Big Ideas: Intergenerational Mobility,” in *CEP CentrePiece*, Winter 2008/9.
- (2011): “Cross-Country Rankings in Intergenerational Mobility: A Comparison of Approaches from Economics and Sociology,” *Journal of Economic Surveys*.
- BLANDEN, J., A. GOODMAN, P. GREGG, AND S. MACHIN (2004): “Changes in Intergenerational Mobility in Britain,” in *Generational Income Mobility in North America and Europe*, pp. 122–46. Cambridge University Press.
- BLANDEN, J., P. GREGG, AND L. MACMILLAN (2012): “Intergenerational Persistence in Income and Social Class: The Effect of Within-Group Inequality,” *Journal of the Royal Statistical Society: Series A*.
- BÖHLMARK, A., AND M. LINDQUIST (2006): “Life-Cycle Variations in the Association between Current and Lifetime Income: Replication and Extension for Sweden,” *Journal of Labor Economics*, 24(4), 879–896.
- BORJAS, G. J. (1999): “The Economic Analysis of Immigration,” in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3 of *Handbook of Labor Economics*, chap. 28, pp. 1697–1760. Elsevier.
- (2003): “The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market*,” *Quarterly journal of Economics*, 118(4), 1335–1374.
- (2009a): “The Analytics of the Wage Effect of Immigration,” NBER Working Papers 14796, National Bureau of Economic Research, Inc.
- (2009b): *Labor Economics*. New York: Irwin/McGraw-Hill, 5th edn.
- BORJAS, G. J., R. B. FREEMAN, AND L. F. KATZ (1996): “Searching for the Effect of Immigration on the Labor Market,” *American Economic Review*, 86(2), 246–51.
- (1997): “How Much Do Immigration and Trade Affect Labor Market Outcomes?,” *Brookings Papers on Economic Activity*, 28(1), 1–90.

- BORJAS, G. J., J. GROGGER, AND G. H. HANSON (2010): “Immigration and the Economic Status of African-American Men,” *Economica*, 77(306), 255–282.
- BOSERUP, S. H., W. KOPCZUK, AND C. T. KREINER (2013): “Intergenerational Wealth Mobility: Evidence from Danish Wealth Records of Three Generations,” Discussion paper, University of Copenhagen.
- BOWLES, S., AND H. GINTIS (2002): “The Inheritance of Inequality,” *Journal of Economic Perspectives*, 16(3), 3–30.
- BRATSBERG, B., AND O. RAAUM (2012): “Immigration and Wages: Evidence from Construction*,” *The Economic Journal*, 122(565), 1177–1205.
- BREEN, R. (2004): “Social Mobility in Europe between 1970 and 2000,” in *Social Mobility in Europe*, ed. by R. Breen, pp. 37–75. Oxford University Press.
- BULMER, M. (2003): *Francis Galton: Pioneer of Heredity and Biometry*. Johns Hopkins University Press.
- CADENA, B. C., AND B. K. KOVAK (2013): “Immigrants Equilibrate Local Labor Markets: Evidence from the Great Recession,” NBER Working Papers 19272, National Bureau of Economic Research.
- CARD, D. (1990): “The Impact of the Mariel Boatlift on the Miami Labor Market,” *Industrial and Labor Relations Review*, pp. 245–257.
- (2001): “Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration,” *Journal of Labor Economics*, 19(1), 22–64.
- (2007): “How Immigration Affects U.S. Cities,” CReAM Discussion Paper Series 0711, Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London.
- (2009): “Immigration and Inequality,” *American Economic Review*, 99(2), 1–21.
- CARD, D., AND J. DiNARDO (2000): “Do Immigrant Inflows Lead to Native Outflows?,” *American Economic Review*, 90(2), 360–367.
- CLARK, G. (2013): “What is the True Rate of Social Mobility? Evidence What is the True Rate of Social Mobility? Evidence from the Information Content of Surnames,” Unpublished working paper, presented at the AEA Annual Meeting, San Diego.

- COLLADO, M. D., I. ORTUÑO-ORTIN, AND A. ROMEU (2013): “Long-run Intergenerational Social Mobility and the Distribution of Surnames,” Discussion paper, Unpublished working paper.
- CONLISK, J. (1969): “An Approach to the Theory of Inequality in the Size Distribution of Income,” *Economic Inquiry*, 7(2), 180–186.
- (1974a): “Can Equalization of Opportunity Reduce Social Mobility?,” *The American Economic Review*, 64(1), pp. 80–90.
- (1974b): “Stability in a Random Coefficient Model,” *International Economic Review*, 15(2), pp. 529–533.
- CORAK, M. (2013): “Inequality from Generation to Generation: The United States in Comparison,” in *The Economics of Inequality, Poverty, and Discrimination in the 21st Century*, ed. by R. Rycroft. ABC-CLIO.
- COUCH, K., AND D. LILLARD (1998): “Sample Selection Rules and the Intergenerational Correlation of Earnings,” *Labour Economics*, 5(3), 313–329.
- CUNHA, F., AND J. HECKMAN (2007): “The Technology of Skill Formation,” *American Economic Review*, 97(2), 31–47.
- DAHL, M., AND T. DELEIRE (2008): “The Association Between Children’s Earnings And Fathers’ Lifetime Earnings: Estimates Using Administrative Data,” Discussion Paper No. 1342-08, Institute for Research on Poverty, University of Wisconsin-Madison.
- DRIBE, M., AND J. HELGERTZ (2013): “Long-term social mobility in Sweden: A three-generation approach,” Discussion paper, mimeo.
- DUSTMANN, C., T. FRATTINI, AND I. PRESTON (2013): “The Effect of Immigration along the Distribution of Wages,” *The Review of Economic Studies*, 80(1).
- DUSTMANN, C., J. LUDSTECK, AND U. SCHÖNBERG (2009): “Revisiting the German Wage Structure,” *The Quarterly Journal of Economics*, 124(2), 843–881.
- DUSTMANN, C., AND I. PRESTON (2012): “Comment: Estimating the Effect of Immigration on Wages,” *Journal of the European Economic Association*, 10(1), 216–223.

- ERIKSON, R., AND J. GOLDTHORPE (1992): *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford University Press, USA.
- (2010): “Has Social Mobility in Britain Decreased? Reconciling Divergent Findings on Income and Class Mobility,” *The British Journal of Sociology*, 61(2), 211–230.
- ERIKSON, R., AND J. O. JONSSON (1996): “Introduction: Explaining Class Inequality in Education: The Swedish Test Case,” in *Can Education Be Equalized? The Swedish Case in Comparative Perspective*, ed. by R. Erikson, and J. Jonsson, pp. 1–64. Boulder, Colorado: Westview Press.
- FITZENBERGER, B., A. OSIKOMINU, AND R. VÖLTER (2006): “Imputation Rules to Improve the Education Variable in the IAB Employment Subsample,” *Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 126(3), 405–436.
- FRIEDMAN, M. (1992): “Do Old Fallacies Ever Die?,” *Journal of Economic Literature*, 30(4), 2129–2132.
- GALOR, O., AND D. TSIDDON (1997): “Technological Progress, Mobility, and Economic Growth,” *American Economic Review*, 87(3), 363–82.
- GOLDBERGER, A. S. (1989): “Economic and Mechanical Models of Intergenerational Transmission,” *The American Economic Review*, 79(3), pp. 504–513.
- GOLDTHORPE, J. H. (2013): “Understanding – and Misunderstanding – Social Mobility in Britain: The Entry of the Economists, the Confusion of Politicians and the Limits of Educational Policy,” *Journal of Social Policy*, 42(03), 431–450.
- GOUSKOVA, E., N. CHITEJI, AND F. STAFFORD (2010): “Estimating the Intergenerational Persistence of Lifetime Earnings with Life Course Matching: Evidence from the PSID,” *Labour Economics*, 17(3), 592–597.
- GRAWE, N. D. (2006): “Lifecycle Bias in Estimates of Intergenerational Earnings Persistence,” *Labour Economics*, 13(5), 551–570.
- GRÖNQVIST, E., B. ÖCKERT, AND J. VLACHOS (2010): “The Intergenerational Transmission of Cognitive and Non-cognitive Abilities,” IFAU Working Paper 2010:12, Institute for Evaluation of Labour Market and Education Policy.

- HAIDER, S., AND G. SOLON (2006): “Life-Cycle Variation in the Association between Current and Lifetime Earnings,” *American Economic Review*, 96(4), 1308–1320.
- HAUSER, R. M. (2010): “Intergenerational Economic Mobility in the United States Measures, Differentials and Trends,” Cde working paper no. 98-12, CDE.
- HAVNES, T., AND M. MOGSTAD (2012): “Is Universal Child Care Leveling the Playing Field?,” Discussion paper, CESifo Working Paper No. 4014.
- HECKMAN, J., J. STIXRUD, AND S. URZUA (2006): “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior,” *Journal of Labor Economics*, 24(3), 411–482.
- HERRNSTEIN, R., AND C. MURRAY (1994): *The Bell Curve: Intelligence and Class Structure in American Life*. Free Press.
- HERTZ, T. (2006): “Understanding Mobility in America,” in *Report For the Center for American Progress*.
- (2007): “Trends in the Intergenerational Elasticity of Family Income in the United States,” *Industrial Relations*, 46(1), 22–50.
- (2008): “A Group-specific Measure of Intergenerational Persistence,” *Economics Letters*, 100(3), 415–417.
- HERTZ, T., T. JAYASUNDERA, P. PIRAINO, S. SELCUK, N. SMITH, AND A. VERASHCHAGINA (2008): “The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends,” *The B.E. Journal of Economic Analysis & Policy*, 7(2), 1–48.
- HODGE, R. (1966): “Occupational Mobility as a Probability Process,” *Demography*, 3(1), 19–34.
- HOLMLUND, H. (2007): “A Researcher’s Guide to the Swedish Compulsory School Reform,” Discussion paper, SOFI.
- (2008): “Intergenerational Mobility and Assortative Mating: Effects of an Educational Reform,” Cee discussion paper 91, Centre for the Economics of Education.

- HOLMLUND, H., M. LINDAHL, AND E. PLUG (2011): “The Causal Effect of Parents’ Schooling on Children’s Schooling: A Comparison of Estimation Methods,” *Journal of Economic Literature*, 49(3), 615–51.
- JÄNTTI, M., AND L. LINDAHL (2012): “On the Variability of Income Within and Across Generations,” *Economics Letters*, 117(1), 165–167.
- JENKINS, S. (1987): “Snapshots versus Movies: ‘Lifecycle biases’ and the Estimation of Intergenerational Earnings Inheritance,” *European Economic Review*, 31(5), 1149–1158.
- JENKINS, S. P. (1982): “Tools for the Analysis of Distributional Models,” *The Manchester School of Economic & Social Studies*, 50(2), 139–50.
- JERRIM, J., AND A. VIGNOLES (2012): “Social Mobility, Regression to the Mean and the Cognitive Development of High Ability Children From Disadvantaged Homes,” *Journal of the Royal Statistical Society: Series A*.
- LEE, C.-I., AND G. SOLON (2009): “Trends in Intergenerational Income Mobility,” *The Review of Economics and Statistics*, 91(4), 766–772.
- LEFGREN, L., M. J. LINDQUIST, AND D. SIMS (2012): “Rich Dad, Smart Dad: Decomposing the Intergenerational Transmission of Income,” *Journal of Political Economy*, 120(2), 268 – 303.
- LEVINE, D. I., AND B. MAZUMDER (2007): “The Growing Importance of Family: Evidence from Brothers’ Earnings,” *Industrial Relations: A Journal of Economy and Society*, 46(1), 7–21.
- LEVY, F., R. MURNANE, AND D. AUTOR (2003): “The Skill Content of Recent Technological Change: An Empirical Exploration,” *The Quarterly Journal of Economics*, 118(4), 1279–1333.
- LINDAHL, M., M. PALME, S. SANDGREN MASSIH, AND A. SJÖGREN (2014): “Dynastic Capital and Long-term Intergenerational Persistence of Earnings and Educational Attainment,” *Journal of Human Resources*, forthcoming.
- LINDQVIST, E., AND R. VESTMAN (2011): “The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment,” *American Economic Journal: Applied Economics*, 3(1), 101–128.

- LONG, J., AND J. FERRIE (2013a): “Grandfathers Matter(ed): Occupational Mobility Across Three Generations in the U.S. and Britain, 1850-1910,” Unpublished working paper, presented at the AEA Annual Meeting, San Diego.
- (2013b): “Intergenerational Occupational Mobility in Great Britain and the United States since 1850,” *American Economic Review*, 103(4), 1109–37.
- MACHIN, S. (2007): “Education Expansion and Intergenerational Mobility in Britain,” in *Schools and the Equal Opportunity Problem*, ed. by L. Woessman, and P. Peterson. MIT Press.
- MANACORDA, M., A. MANNING, AND J. WADSWORTH (2012): “The Impact Of Immigration On The Structure Of Wages: Theory And Evidence From Britain,” *Journal of the European Economic Association*, 10(1), 120–151.
- MANKIW, G. (2006): “Rich Dad, Poor Dad,” in *Greg Mankiw’s Blog*, published on April 27, 2006. <http://gregmankiw.blogspot.co.uk/2006/04/rich-dad-poor-dad.html>.
- MARE, R. (2011): “A Multigenerational View of Inequality,” *Demography*, 48(1), 1–23.
- MAZUMDER, B. (2005): “Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data,” *Review of Economics and Statistics*, 87(2), 235–255.
- MEGHIR, C., AND M. PALME (2005): “Educational Reform, Ability, and Family Background,” *American Economic Review*, 95(1), 414–424.
- MEGHIR, C., M. PALME, AND M. SCHNABEL (2011): “The Effect of Education Policy on Crime: An Intergenerational Perspective,” IZA Discussion Papers 6142, Institute for the Study of Labor (IZA).
- MOCETTI, S. (2007): “Intergenerational Earnings Mobility in Italy,” *The B.E. Journal of Economic Analysis & Policy*, 7(2), Article 5.
- NESSELROADE, J. R., S. M. STIGLER, AND P. B. BALTES (1980): “Regression Toward the Mean and the Study of Change,” *Psychological Bulletin*, 88(3).
- NICOLETTI, C., AND J. ERMISCH (2007): “Intergenerational Earnings Mobility: Changes across Cohorts in Britain,” *The B.E. Journal of Economic Analysis & Policy*, 7(2), Article 9.

- NILSEN, Ø., K. VAAGE, A. AAKVIK, AND K. JACOBSEN (2012): “Intergenerational Earnings Mobility Revisited: Estimates Based on Lifetime Earnings,” *Scandinavian Journal of Economics*, 114(1), 1–23.
- NYBOM, M., AND J. STUHLER (2011): “Heterogeneous Income Profiles and Life-Cycle Bias in Intergenerational Mobility Estimation,” IZA Discussion Papers 5697, Institute for the Study of Labor (IZA).
- OLIVETTI, C., D. PASERMAN, AND L. SALISBURY (2014): “Intergenerational Mobility Across Three Generations in the 19th Century: Evidence from the US Census,” Discussion paper, Unpublished working paper.
- OTTAVIANO, G. I. P., AND G. PERI (2012): “Rethinking The Effect Of Immigration On Wages,” *Journal of the European Economic Association*, 10(1), 152–197.
- PEKKARINEN, T., R. UUSITALO, AND S. KERR (2009): “School Tracking and Intergenerational Income Mobility: Evidence from the Finnish Comprehensive School Reform,” *Journal of Public Economics*, 93(7-8), 965–973.
- PIKETTY, T. (2000): “Theories of Persistent Inequality and Intergenerational Mobility,” in *Handbook of Income Distribution*, ed. by A. B. Atkinson, and F. Bourguignon, vol. 1, pp. 429 – 476. Elsevier.
- PIRAINO, P. (2007): “Comparable Estimates of Intergenerational Income Mobility in Italy,” *The B.E. Journal of Economic Analysis & Policy*, 7(2), Article 1.
- RAAUM, O., B. BRATSBERG, K. RØED, E. ÖSTERBACKA, T. ERIKSSON, M. JÄNTTI, AND R. NAYLOR (2007): “Marital Sorting, Household Labor Supply, and Intergenerational Earnings Mobility across Countries,” *The B.E. Journal of Economic Analysis & Policy*, 7(2), Article 7.
- REDDING, S., AND D. STURM (2008): “The Costs of Remoteness: Evidence From German Division and Reunification,” *American Economic Review*, 98(5), 1766–1797.
- SMITH, C. L. (2012): “The Impact of Low-Skilled Immigration on the Youth Labor Market,” *Journal of Labor Economics*, 30(1), pp. 55–89.
- SOLON, G. (1999): “Intergenerational Mobility in the Labor Market,” in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3A, chap. 29, pp. 1761–1800. Elsevier.

- (2004): “A Model of Intergenerational Mobility Variation over Time and Place,” *Generational Income Mobility in North America and Europe*, pp. 38–47.
- (2013): “Theoretical Models of Inequality Transmission across Multiple Generations,” NBER Working Papers 18790, National Bureau of Economic Research, Inc.
- VOGEL, T. (2006): “Reassessing Intergenerational Mobility in Germany and the United States: The Impact of Differences in Lifecycle Earnings Patterns,” SFB 649 Discussion Papers 2006-055, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany.
- WARREN, J. R., AND R. M. HAUSER (1997): “Social Stratification across Three Generations: New Evidence from the Wisconsin Longitudinal Study,” *American Sociological Review*, 62(4), pp. 561–572.
- ZAX, J., AND D. REES (2002): “IQ, Academic Performance, Environment, and Earnings,” *Review of Economics and Statistics*, 84(4), 600–616.
- ZYLBERBERG, Y. (2013): “Dynastic Income Mobility,” Discussion paper, CREi.